

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 02/097059 A2(51) International Patent Classification⁷: C12N

(21) International Application Number: PCT/US02/17452

(22) International Filing Date: 30 May 2002 (30.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/294,758 30 May 2001 (30.05.2001) US
60/366,891 21 March 2002 (21.03.2002) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US 60/294,758 (CIP)
Filed on 30 May 2001 (30.05.2001)
US 60/366,891 (CIP)
Filed on 21 March 2002 (21.03.2002)

(71) Applicant (for all designated States except US): CHROMOS MOLECULAR SYSTEMS, INC. [CA/CA]; 8081 Lougheed Highway, Burnaby, B.C. V5A 1W9 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): PERKINS, Edward [US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6 (CA). PEREZ, Carl [US/CA]; 1201-7680 Granville Avenue, Richmond, B.C. V6Y 4B9 (CA). LINDENBAUM, Michael [CA/CA]; 252 Finnigan Street, Coquitlam, B.C. V3K 5J7 (CA). GREENE, Amy [US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6 (CA). LEUNG, Josephine [CA/CA]; 711 Ebert Avenue, Coquitlam, B.C. V3J 7P8 (CA). FLEMING, Elena [CA/CA]; 248 E 18th, North Vancouver, B.C. V7L 2X6 (CA). STEWART, Sandra [CA/CA]; 2618 Oxford Street, Vancouver, B.C. V5K 1N3 (CA). SHELLARD, Joan [CA/CA]; #215-1345 West 15th Avenue, Vancouver, B.C. V6H 3R3 (CA).

(74) Agents: SEIDMAN, Stephanie, L. et al.; Heller Ehrman White & McAuliffe LLP, 7th floor, 4350 La Jolla Village Drive, San Diego, CA 92122-1246 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,

MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: CHROMOSOME-BASED PLATFORMS

(57) Abstract: Artificial chromosomes, including *Aces*, that have been engineered to contain available sites for site-specific, recombination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational engineering of the chromosome for a variety of applications.

WO 02/097059 A2

CHROMOSOME-BASED PLATFORMS**RELATED APPLICATIONS**

Benefit of priority to U.S. provisional application Serial No. 60/294,758, filed May 30, 2001, to Perkins, *et al.*, entitled "CHROMOSOME-BASED PLATFORMS" and to U.S. provisional application Serial No. 60/366,891, filed March 21, 2002, to Perkins, *et al.*, entitled

5 "CHROMOSOME-BASED PLATFORMS" is claimed. Where permitted, the subject matter of which are herein incorporated by reference in their entirety.

This application is related to Provisional Application No. 60/294,687, filed May 30, 2001, by CARL PEREZ AND STEVEN

10 FABIJANSKI entitled *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING PLANT ARTIFICIAL CHROMOSOMES* and to U.S. Provisional Application No. 60/296,329, filed June 4, 2001, by CARL PEREZ AND STEVEN FABIJANSKI entitled

15 *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING PLANT ARTIFICIAL CHROMOSOMES*. This application also is related to U.S. Provisional Application No. 60/294,758, filed May 30, 2001, by EDWARD PERKINS *et al.*, entitled *CHROMOSOME-BASED PLATFORMS* and to U.S. Provisional Application No. 60/366,891, filed March 21, 2002, by EDWARD PERKINS *et al.*, entitled

20 *CHROMOSOME-BASED PLATFORMS*. This application is also related to U.S. application Serial Nos. (attorney dkt nos. 24601-419 and 419PC), filed on the same day herewith, entitled *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS OF PREPARING PLANT ARTIFICIAL CHROMOSOMES* to Perez *et al.*.

25 This application is related to U.S. application Serial No. 08/695,191, filed August 7, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND*

-2-

METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES, now U.S. Patent No. 6,025,155. This application is also related to U.S. application Serial No. 08/682,080, filed July 15, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF*
5 *AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*, now U.S. Patent No. 6,077,697. This application is also related U.S. application Serial No. 08/629,822, filed April 10, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING*
10 *ARTIFICIAL CHROMOSOMES* (now abandoned), and is also related to copending U.S. application Serial No. 09/096,648, filed June 12, 1998, by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* and to U.S. application Serial No.
15 09/835,682, April 10, 1997 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* (now abandoned). This application is also related to copending U.S. application Serial No. 09/724,726, filed November 28, 2000, U.S. application Serial
20 No. 09/724,872, filed November 28, 2000, U.S. application Serial No. 09/724,693, filed November 28, 2000, U.S. application Serial No. 09/799,462, filed March 5, 2001, U.S. application Serial No. 09/836,911, filed April 17, 2001, and U.S. application Serial No. 10/125,767, filed April 17, 2002, each of which is by GYULA
25 HADLACZKY and ALADAR SZALAY, and is entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*. This application is also related to International PCT application No. WO 97/40183. Where permitted the

-3-

subject matter of each of these provisional applications, international applications, and applications is incorporated by reference in its entirety.

FIELD OF INVENTION

Artificial chromosomes, including *ACes*, that have been engineered
5 to contain available sites for site-specific, recombination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational engineering of the chromosome.

BACKGROUND

Artificial chromosomes

10 A variety of artificial chromosomes for use in plants and animals, particularly higher plants and animals are available. In particular, U.S. Patent Nos. 6,025,155 and 6,077,697 provide heterochromatic artificial chromosomes designated therein as satellite artificial chromosomes (SATACs) and now designated artificial chromosome expression systems
15 (*ACes*). These chromosomes are prepared by introducing heterologous DNA into a selected plant or animal cell under conditions that result in integration into a region of the chromosome that leads to an amplification event resulting in production of a dicentric chromosome. Subsequent treatment and growth of cells with dicentric chromosomes, including
20 further amplifications, ultimately results in the artificial chromosomes provided therein. In order to introduce a desired heterologous gene (or a plurality of heterologous genes) into the artificial chromosome, the process is repeated introducing the desired heterologous genes and nucleic acids in the initial targeting step. This process is time consuming
25 and tedious. Hence, more tractable and efficient methods for introducing heterologous nucleic acid molecules into artificial chromosomes, particularly *ACes*, are needed.

-4-

Therefore, it is an object herein to provide engineered artificial chromosomes that permit tractable, efficient and rational engineering of artificial chromosomes.

SUMMARY OF THE INVENTION

5 Provided herein are artificial chromosomes that permit tractable, efficient and rational engineering thereof. In particular, the artificial chromosomes provided herein contain one or a plurality of loci (sites) for site-specific, recombination-directed integration of DNA. Thus, provided herein are platform artificial chromosome expression systems ("platform
10 *ACes*") containing single or multiple site-specific, recombination sites. The artificial chromosomes and *ACes* artificial chromosomes include plant and animal chromosomes. Any recombinase system that effects site-specific recombination is contemplated for use herein.

 In one embodiment, chromosomes, including platform *ACes*, are
15 provided that contain one or more lambda *att* sites designed for recombination-directed integration in the presence of lambda integrase, and that are mutated so that they do not require additional factors. Methods for preparing such chromosomes, vectors for use in the methods, and uses of the resulting chromosomes are also provided.

20 Platform *ACes* containing the recombination site(s) and methods for introducing heterologous nucleic acid into such sites and vectors therefor, are provided.

 Also provided herein is a bacteriophage lambda (λ) integrase site-specific recombination system.

25 Methods using recombinase mediated recombination target gene expression vectors and/or genes for insertion thereof into platform chromosomes and the resulting chromosomes are provided.

 Combinations and kits containing the combinations of vectors encoding a recombinase and integrase and primers for introduction of the

-5-

site recognized thereby are also provided. The kits optionally include instructions for performing site-directed integration or preparation of *ACes* containing such sites.

Also provided herein are mammalian and plant cells comprising the
5 artificial chromosomes and *ACes* described herein. The cells can be nuclear donor cells, stem cells, such as a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell or an embryonic stem cell.

Also provide is a lambda-intR mutain comprising a glutamic acid to arginine change at position 174 of wild-type lambda-integrase3. Also
10 provided are transgenic animals and methods for producing a transgenic animal, comprising introducing a *ACes* into an embryonic cell, such as a stem cell or embryo. The *ACes* can comprise heterologous nucleic acid that encodes a therapeutic product. The transgenic animal can be a fish, insect, reptile, amphibians, arachnid or mammal. In certain embodiments,
15 the *ACes* is introduced by cell fusion, lipid-mediated transfection by a carrier system, microinjection, microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer.

The platform *ACes*, including plant and animal *ACes*, such as MACs, provided herein can be introduced into cells, such as, but not
20 limited to, animal cells, including mammalian cells, and into plant cells. Hence plant cells that contain platform MACs, animal cells that contain platform PACs and other combinations of cells and platform *ACes* are provided.

DESCRIPTION OF FIGURES

25 FIGURE 1 provides a diagram depicting creation of an exemplary *ACes* artificial chromosome prepared using methods detailed in U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183. In this exemplified embodiment, the nucleic acid is targeted to an acrocentric chromosome in an animal or plant, and the

-6-

heterologous nucleic acid includes a sequence-specific recombination site and marker genes.

FIGURE 2 provides a map of pWEPuro9K, which is a targeting vector derived from the vector pWE15 (GenBank Accession # X65279; SEQ ID No. 31). Plasmid pWE15 was modified by replacing the *Sal*I (Klenow filled)/*Sma*I neomycin resistance encoding fragment with the *Pvu*II/*Bam*HI (Klenow filled) puromycin resistance-encoding fragment (isolated from plasmid pPUR, Clontech Laboratories, Inc., Palo Alto, CA; GenBank Accession no. U07648; SEQ ID No. 30) resulting in plasmid pWEPuro. Subsequently a 9 Kb *Not*I fragment from the plasmid pFK161 (see Example 1, see, also Csonka *et al.* (2000) *Journal of Cell Science* 113:3207-32161; and SEQ ID NO: 118), containing a portion of the mouse rDNA region, was cloned into the *Not*I site of pWEPuro resulting in plasmid pWEPuro9K.

FIGURE 3 depicts construction of an ACes platform chromosome with a single recombination site, such as loxP sites or an *att*P or *att*B site. This platform ACes chromosome is an exemplary artificial chromosome with a single recombination site.

FIGURE 4 provides a map of plasmid pSV40-193attPsensePur.

FIGURE 5 depicts a method for formation of a chromosome platform with multiple recombination integration sites, such as *att*P sites.

FIGURE 6 sets forth the sequences of the core region of *att*P, *att*B, *att*L and *att*R (SEQ ID Nos. 33-36).

FIGURE 7 depicts insertional recombination of a vector encoding a marker gene, DsRed and an *att*B site with an artificial chromosome containing an *att*P site.

FIGURE 8 provides a map of plasmid pCXLamIntR (SEQ ID NO: 112), which includes the Lambda integrase (E174R)-encoding nucleic acid.

-7-

FIGURE 9 diagrammatically summarizes the platform technology; marker 1 permits selection of the artificial chromosomes containing the integration site; marker 2, which is promoterless in the target gene expression vector, permits selection of recombinants. Upon
5 recombination with the platform marker 2 is expressed under the control of a promoter resident on the platform.

FIGURE 10 provides the vector map for the plasmid p18attBZEO-5'6XHS4eGFP (SEQ ID NO: 116).

FIGURE 11 provides the vector map for the plasmid p18attBZEO-
10 3'6XHS4eGFP (SEQ ID NO: 115).

FIGURE 12 provides the vector map for the plasmid p18attBZEO-(6XHS4)2eGFP (SEQ ID NO: 110).

FIGURES 13 AND 14 depict the integration of a PCR product by site-specific recombination as set forth in Example 8.

15 FIGURE 15 provides the vector map for the plasmid pPACrDNA as set forth in Example 9.A.

DETAILED DESCRIPTION OF THE INVENTION

A. DEFINITIONS

Unless defined otherwise, all technical and scientific terms used
20 herein have the same meaning as is commonly understood by one of skill in the art to which the invention(s) belong. All patents, patent applications, published applications and publications, Genbank sequences, websites and other published materials referred to throughout the entire disclosure herein, unless noted otherwise, are incorporated by reference
25 in their entirety. Where reference is made to a URL or other such identifier or address, it is understood that such identifiers can change and particular information on the internet can come and go, but equivalent information can be found by searching the internet. Reference thereto evidences the availability and public dissemination of such information.

-8-

As used herein, nucleic acid refers to single-stranded and/or double-stranded polynucleotides, such as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), as well as analogs or derivatives of either RNA or DNA. Also included in the term "nucleic acid" are analogs of nucleic acids such as peptide nucleic acid (PNA), phosphorothioate DNA, and other such analogs and derivatives. When referring to probes or primers, optionally labeled, with a detectable label, such as a fluorescent or radiolabel, single-stranded molecules are contemplated. Such molecules are typically of a length such that they are statistically unique and of low copy number (typically less than 5, preferably less than 3) for probing or priming a library. Generally a probe or primer contains at least 14, 16 or 30 contiguous nucleotides of sequence complementary to or identical to a gene of interest. Probes and primers can be 10, 20, 30, 50, 100 or more nucleotides long.

As used herein, DNA is meant to include all types and sizes of DNA molecules including cDNA, plasmids and DNA including modified nucleotides and nucleotide analogs.

As used herein, nucleotides include nucleoside mono-, di-, and triphosphates. Nucleotides also include modified-nucleotides, such as, but are not limited to, phosphorothioate nucleotides and deazapurine nucleotides and other nucleotide analogs.

As used herein, heterologous or foreign DNA and RNA are used interchangeably and refer to DNA or RNA that does not occur naturally as part of the genome in which it is present or which is found in a location or locations and/or in amounts in a genome or cell that differ from that in which it occurs in nature. Heterologous nucleic acid is generally not endogenous to the cell into which it is introduced, but has been obtained from another cell or prepared synthetically. Generally, although not necessarily, such nucleic acid encodes RNA and proteins that are not

-9-

normally produced by the cell in which it is expressed. Any DNA or RNA that one of skill in the art would recognize or consider as heterologous or foreign to the cell in which it is expressed is herein encompassed by heterologous DNA. Heterologous DNA and RNA may also encode RNA or proteins that mediate or alter expression of endogenous DNA by affecting transcription, translation, or other regulatable biochemical processes.

Examples of heterologous DNA include, but are not limited to, DNA that encodes a gene product or gene product(s) of interest, introduced for purposes of modification of the endogenous genes or for production of an encoded protein. For example, a heterologous or foreign gene may be isolated from a different species than that of the host genome, or alternatively, may be isolated from the host genome but operably linked to one or more regulatory regions which differ from those found in the unaltered, native gene. Other examples of heterologous DNA include, but are not limited to, DNA that encodes traceable marker proteins, such as a protein that confers traits including, but not limited to, herbicide, insect, or disease resistance; traits, including, but not limited to, oil quality or carbohydrate composition. Antibodies that are encoded by heterologous DNA may be secreted or expressed on the surface of the cell in which the heterologous DNA has been introduced.

As used herein, operative linkage or operative association, or grammatical variations thereof, of heterologous DNA to regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences refers to the relationship between such DNA and such sequences of nucleotides. For example, operative linkage of heterologous DNA to a promoter refers to the physical relationship between the DNA and the promoter such that the transcription of such DNA is initiated from the

-10-

promoter by an RNA polymerase that specifically recognizes, binds to and transcribes the DNA.

In order to optimize expression and/or *in vitro* transcription, it may be necessary to remove, add or alter 5' untranslated portions of the
5 clones to eliminate extra, potential inappropriate alternative translation initiation (*i.e.*, start) codons or other sequences that may interfere with or reduce expression, either at the level of transcription or translation. Alternatively, consensus ribosome binding sites (see, *e.g.*, Kozak (1991) *J. Biol. Chem.* 266:19867-19870) can be inserted immediately 5' of the
10 start codon and may enhance expression.

As used herein, a sequence complementary to at least a portion of an RNA, with reference to antisense oligonucleotides, means a sequence having sufficient complementarity to be able to hybridize with the RNA, preferably under moderate or high stringency conditions, forming a stable
15 duplex. The ability to hybridize depends on the degree of complementarity and the length of the antisense nucleic acid. The longer the hybridizing nucleic acid, the more base mismatches it can contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard
20 procedures to determine the melting point of the hybridized complex.

As used herein, regulatory molecule refers to a polymer of deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) or a polypeptide that is capable of enhancing or inhibiting expression of a gene.

As used herein, recognition sequences are particular sequences of
25 nucleotides that a protein, DNA, or RNA molecule, or combinations thereof, (such as, but not limited to, a restriction endonuclease, a modification methylase and a recombinase) recognizes and binds. For example, a recognition sequence for Cre recombinase (see, *e.g.*, SEQ ID NO:58) is a 34 base pair sequence containing two 13 base pair inverted

-11-

repeats (serving as the recombinase binding sites) flanking an 8 base pair core and designated loxP (see, *e.g.*, Sauer (1994) *Current Opinion in Biotechnology* 5:521-527). Other examples of recognition sequences, include, but are not limited to, attB and attP, attR and attL and others
5 (see, *e.g.*, SEQ ID Nos. 8, 41-56 and 72), that are recognized by the recombinase enzyme Integrase (see, SEQ ID Nos. 37 and 38 for the nucleotide and encoded amino acid sequences of an exemplary lambda phage integrase).

The recombination site designated attB is an approximately 33 base
10 pair sequence containing two 9 base pair core-type Int binding sites and a 7 base pair overlap region; attP (SEQ ID No. 72) is an approximately 240 base pair sequence containing core-type Int binding sites and arm-type Int binding sites as well as sites for auxiliary proteins IHF, FIS, and Xis (see, *e.g.*, Landy (1993) *Current Opinion in Biotechnology* 3:699-707) see,
15 *e.g.*, SEQ ID Nos. 8 and 72).

As used herein, a recombinase is an enzyme that catalyzes the exchange of DNA segments at specific recombination sites. An integrase herein refers to a recombinase that is a member of the lambda (λ) integrase family.

20 As used herein, recombination proteins include excisive proteins, integrative proteins, enzymes, co-factors and associated proteins that are involved in recombination reactions using one or more recombination sites (see, Landy (1993) *Current Opinion in Biotechnology* 3:699-707). The recombination proteins used herein can be delivered to a cell via an
25 expression cassette on an appropriate vector, such as a plasmid, and the like. In other embodiments, the recombination proteins can be delivered to a cell in protein form in the same reaction mixture used to deliver the desired nucleic acid, such as a platform ACes, donor target vectors, and the like.

-12-

As used herein the expression "lox site" means a sequence of nucleotides at which the gene product of the cre gene, referred to herein as Cre, can catalyze a site-specific recombination event. A LoxP site is a 34 base pair nucleotide sequence from bacteriophage P1 (see, 5 *e.g.*, Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398-3402). The LoxP site contains two 13 base pair inverted repeats separated by an 8 base pair spacer region as follows: (SEQ ID NO. 57):

ATAACTTCGTATA ATGTATGC TATACGAAGTTAT

E. coli DH5Δlac and yeast strain BSY23 transformed with plasmid pBS44 10 carrying two loxP sites connected with a LEU2 gene are available from the American Type Culture Collection (ATCC) under accession numbers ATCC 53254 and ATCC 20773, respectively. The lox sites can be isolated from plasmid pBS44 with restriction enzymes *EcoRI* and *Sall*, or *XhoI* and *BamHI*. In addition, a preselected DNA segment can be inserted 15 into pBS44 at either the *Sall* or *BamHI* restriction enzyme sites. Other lox sites include, but are not limited to, LoxB, LoxL, LoxC2 and LoxR sites, which are nucleotide sequences isolated from *E. coli* (see, *e.g.*, Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398). Lox sites can also be produced by a variety of synthetic techniques (see, *e.g.*, Ito *et al.* (1982) 20 *Nuc. Acid Res.* 10:1755 and Ogilvie *et al.* (1981) *Science* 270:270).

As used herein, the expression "cre gene" means a sequence of nucleotides that encodes a gene product that effects site-specific recombination of DNA in eukaryotic cells at lox sites. One cre gene can be isolated from bacteriophage P1 (see, *e.g.*, Abremski *et al.* (1983) *Cell* 25 32:1301-1311). *E. coli* DH1 and yeast strain BSY90 transformed with plasmid pBS39 carrying a cre gene isolated from bacteriophage P1 and a GAL1 regulatory nucleotide sequence are available from the American Type Culture Collection (ATCC) under accession numbers ATCC 53255

-13-

and ATCC 20772, respectively. The cre gene can be isolated from plasmid pBS39 with restriction enzymes *XhoI* and *Sall*.

As used herein, site-specific recombination refers to site-specific recombination that is effected between two specific sites on a single
5 nucleic acid molecule or between two different molecules that requires the presence of an exogenous protein, such as an integrase or recombinase.

For example, Cre-lox site-specific recombination can include the following three events:

- 10 a. deletion of a pre-selected DNA segment flanked by lox sites;
- b. inversion of the nucleotide sequence of a pre-selected DNA segment flanked by lox sites; and
- c. reciprocal exchange of DNA segments proximate to
15 lox sites located on different DNA molecules.

This reciprocal exchange of DNA segments can result in an integration event if one or both of the DNA molecules are circular. DNA segment refers to a linear fragment of single- or double-stranded deoxyribonucleic acid (DNA), which can be derived from any source.

20 Since the lox site is an asymmetrical nucleotide sequence, two lox sites on the same DNA molecule can have the same or opposite orientations with respect to each other. Recombination between lox sites in the same orientation results in a deletion of the DNA segment located between the two lox sites and a connection between the resulting ends of the original
25 DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule result in an inversion of the nucleotide sequence of the DNA segment located between the two lox

-14-

sites. In addition, reciprocal exchange of DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the gene product of the cre gene. Thus, the Cre-lox system can be used to specifically delete, invert, or
5 insert DNA. The precise event is controlled by the orientation of lox DNA sequences, in *cis* the lox sequences direct the *Cre* recombinase to either delete (lox sequences in direct orientation) or invert (lox sequences in inverted orientation) DNA flanked by the sequences, while *in trans* the lox sequences can direct a homologous recombination event resulting in the
10 insertion of a recombinant DNA.

As used herein, a chromosome is a nucleic acid molecule, and associated proteins, that is capable of replication and segregation within a cell upon cell division. Typically, a chromosome contains a centromeric region, replication origins, telomeric regions and a region of nucleic acid
15 between the centromeric and telomeric regions.

As used herein, a centromere is any nucleic acid sequence that confers an ability to segregate to daughter cells through cell division. A centromere may confer stable segregation of a nucleic acid sequence, including an artificial chromosome containing the centromere, through
20 mitotic or meiotic divisions, including through both mitotic and meiotic divisions. A particular centromere is not necessarily derived from the same species in which it is introduced, but has the ability to promote DNA segregation in cells of that species.

As used herein, euchromatin and heterochromatin have their
25 recognized meanings. Euchromatin refers to chromatin that stains diffusely and that typically contains genes, and heterochromatin refers to chromatin that remains unusually condensed and that has been thought to be transcriptionally inactive. Highly repetitive DNA sequences (satellite DNA) are usually located in regions of the heterochromatin surrounding

-15-

the centromere (pericentric or pericentromeric heterochromatin).

Constitutive heterochromatin refers to heterochromatin that contains the highly repetitive DNA which is constitutively condensed and genetically inactive.

5 As used herein, an acrocentric chromosome refers to a chromosome with arms of unequal length.

As used herein, endogenous chromosomes refer to genomic chromosomes as found in a cell prior to generation or introduction of an artificial chromosome.

10 As used herein, artificial chromosomes are nucleic acid molecules, typically DNA, that stably replicate and segregate alongside endogenous chromosomes in cells and have the capacity to accommodate and express heterologous genes contained therein. It has the capacity to act as a gene delivery vehicle by accommodating and expressing foreign genes
15 contained therein. A mammalian artificial chromosome (MAC) refers to chromosomes that have an active mammalian centromere(s). Plant artificial chromosomes, insect artificial chromosomes and avian artificial chromosomes refer to chromosomes that include centromeres that function in plant, insect and avian cells, respectively. A human artificial
20 chromosome (HAC) refers to chromosomes that include centromeres that function in human cells. For exemplary artificial chromosomes, see, *e.g.*, U.S. Patent Nos. 6,025,155; 6,077,697; 5,288,625; 5,712,134; 5,695,967; 5,869,294; 5,891,691 and 5,721,118 and published International PCT application Nos, WO 97/40183 and WO 98/08964.

25 Artificial chromosomes include those that are predominantly heterochromatic (formerly referred to as satellite artificial chromosomes (SATACs); see, *e.g.*, U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183), minichromosomes that contain a *de novo* centromere (see, U.S. Patent

-16-

Nos. 5,712,134, 5,891,691 and 5,288,625), artificial chromosomes predominantly made up of repeating nucleic acid units and that contain substantially equivalent amounts of euchromatic and heterochromatic DNA and *in vitro* assembled artificial chromosomes (see, copending U.S. 5 provisional application Serial No. 60/294,687, filed on May 30, 2001).

As used herein, the term "satellite DNA-based artificial chromosome (SATAC)" is interchangeable with the term "artificial chromosome expression system (ACes)". These artificial chromosomes (ACes) include those that are substantially all neutral non-coding 10 sequences (heterochromatin) except for foreign heterologous, typically gene-encoding nucleic acid, that is interspersed within the heterochromatin for the expression therein (see U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183), or that is in a single locus as provided herein. Also included 15 are ACes that may include euchromatin and that result from the process described in U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183 and outlined herein. The delineating structural feature is the presence of repeating units, that are generally predominantly heterochromatin. The precise structure of the ACes will 20 depend upon the structure of the chromosome in which the initial amplification event occurs; all share the common feature of including a defined pattern of repeating units. Generally ACes have more heterochromatin than euchromatin. Foreign nucleic acid molecules (heterologous genes) contained in these artificial chromosome expression 25 systems can include any nucleic acid whose expression is of interest in a particular host cell. Such foreign nucleic acid molecules, include, but are not limited to, nucleic acid that encodes traceable marker proteins (reporter genes), such as fluorescent proteins, such as green, blue or red fluorescent proteins (GFP, BFP and RFP, respectively), other reporter

-17-

genes, such as β -galactosidase and proteins that confer drug resistance, such as a gene encoding hygromycin-resistance. Other examples of heterologous nucleic acid molecules include, but are not limited to, DNA that encodes therapeutically effective substances, such as anti-cancer
5 agents, enzymes and hormones, DNA that encodes other types of proteins, such as antibodies, and DNA that encodes RNA molecules (such as antisense or siRNA molecules) that are not translated into proteins.

As used herein, an artificial chromosome platform, also referred to
10 herein as a "platform ACes" or "ACes platform", refers to an artificial chromosome that has been engineered to include one or more sites for site-specific, recombination-directed integration. In particular, ACes that are so-engineered are provided. Any sites, including but not limited to any described herein, that are suitable for such integration are
15 contemplated. Plant and animal platform ACes are provided. Among the ACes contemplated herein are those that are predominantly heterochromatic (formerly referred to as satellite artificial chromosomes (SATACs); see, *e.g.*, U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183), artificial
20 chromosomes predominantly made up of repeating nucleic acid units and that contain substantially equivalent amounts of euchromatic and heterochromatic DNA resulting from an amplification event depicted in the referenced patent and herein. Included among the ACes for use in generating platforms, are artificial chromosomes that introduce and
25 express heterologous nucleic acids in plants (see, copending U.S. provisional application Serial No. 60/294,687, filed on May 30, 2001). These include artificial chromosomes that have a centromere derived from a plant, and, also, artificial chromosomes that have centromeres that may be derived from other organisms but that function in plants.

-18-

As used herein a "reporter ACes" refers to a an ACes that comprises one or a plurality of reporter constructs, where the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

5 As used herein, amplification, with reference to DNA, is a process in which segments of DNA are duplicated to yield two or multiple copies of substantially similar or identical or nearly identical DNA segments that are typically joined as substantially tandem or successive repeats or inverted repeats.

10 As used herein, amplification-based artificial chromosomes are artificial chromosomes derived from natural or endogenous chromosomes by virtue of an amplification event, such as one initiated by introduction of heterologous nucleic acid into rDNA in a chromosome. As a result of such an event, chromosomes and fragments thereof exhibiting segmented
15 or repeating patterns arise. Artificial chromosomes can be formed from these chromosomes and fragments. Hence, amplification-based artificial chromosomes refer to engineered chromosomes that exhibit an ordered segmentation that is not observed in naturally occurring chromosomes and that distinguishes them from naturally occurring chromosomes. The
20 segmentation, which can be visualized using a variety of chromosome analysis techniques known to those of skill in the art, correlates with the structure of these artificial chromosomes. In addition to containing one or more centromeres, the amplification-based artificial chromosomes, throughout the region or regions of segmentation are predominantly made
25 up of nucleic acid units also referred to as "amplicons", that is (are) repeated in the region and that have a similar gross structure. Repeats of an amplicon tend to be of similar size and share some common nucleic acid sequences. For example, each repeat of an amplicon may contain a replication site involved in amplification of chromosome segments and/or

-19-

some heterologous nucleic acid that was utilized in the initial production of the artificial chromosome. Typically, the repeating units are substantially similar in nucleic acid composition and may be nearly identical.

5 The amplification-based artificial chromosomes differ depending on the chromosomal region that has undergone amplification in the process of artificial chromosome formation. The structures of the resulting chromosomes can vary depending upon the initiating event and/or the conditions under which the heterologous nucleic acid is introduced,
10 including modification to the endogenous chromosomes. For example, in some of the artificial chromosomes provided herein, the region or regions of segmentation may be made up predominantly of heterochromatic DNA. In other artificial chromosomes provided herein, the region or regions of segmentation may be made up predominantly of euchromatic DNA or may
15 be made up of similar amounts of heterochromatic and euchromatic DNA.

As used herein an amplicon is a repeated nucleic acid unit. In some of the artificial chromosomes described herein, an amplicon may contain a set of inverted repeats of a megareplicon. A megareplicon represents a higher order replication unit. For example, with reference to
20 some of the predominantly heterochromatic artificial chromosomes, the megareplicon can contain a set of tandem DNA blocks (*e.g.*, ~7.5 Mb DNA blocks) each containing satellite DNA flanked by non-satellite DNA or may be made up of substantially rDNA. Contained within the megareplicon is a primary replication site, referred to as the
25 megareplicator, which may be involved in organizing and facilitating replication of the pericentric heterochromatin and possibly the centromeres. Within the megareplicon there may be smaller (*e.g.*, 50-300 kb) secondary replicons.

-20-

In artificial chromosomes, such as those provided U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183, the megareplicon is defined by two tandem blocks (~7.5 Mb DNA blocks in the chromosomes provided therein). Within each artificial
5 chromosome or among a population thereof, each amplicon has the same gross structure but may contain sequence variations. Such variations will arise as a result of movement of mobile genetic elements, deletions or insertions or mutations that arise, particularly in culture. Such variation does not affect the use of the artificial chromosomes or their overall
10 structure as described herein.

As used herein, amplifiable, when used in reference to a chromosome, particularly the method of generating artificial chromosomes provided herein, refers to a region of a chromosome that is prone to amplification. Amplification typically occurs during replication and other
15 cellular events involving recombination (*e.g.*, DNA repair). Such regions include regions of the chromosome that contain tandem repeats, such as satellite DNA, rDNA, and other such sequences.

As used herein, a dicentric chromosome is a chromosome that contains two centromeres. A multicentric chromosome contains more
20 than two centromeres.

As used herein, a formerly dicentric chromosome is a chromosome that is produced when a dicentric chromosome fragments and acquires new telomeres so that two chromosomes, each having one of the centromeres, are produced. Each of the fragments is a replicable
25 chromosome. If one of the chromosomes undergoes amplification of primarily euchromatic DNA to produce a fully functional chromosome that is predominantly (at least more than 50%) euchromatin, it is a minichromosome. The remaining chromosome is a formerly dicentric chromosome. If one of the chromosomes undergoes amplification,

-21-

whereby heterochromatin (such as, for example, satellite DNA) is amplified and a euchromatic portion (such as, for example, an arm) remains, it is referred to as a sausage chromosome. A chromosome that is substantially all heterochromatin, except for portions of heterologous DNA, is called a predominantly heterochromatic artificial chromosome. Predominantly heterochromatic artificial chromosomes can be produced from other partially heterochromatic artificial chromosomes by culturing the cell containing such chromosomes under conditions such as BrdU treatment that destabilize the chromosome and/or growth under selective conditions so that a predominantly heterochromatic artificial chromosome is produced. For purposes herein, it is understood that the artificial chromosomes may not necessarily be produced in multiple steps, but may appear after the initial introduction of the heterologous DNA. Typically, artificial chromosomes appear after about 5 to about 60, or about 5 to about 55, or about 10 to about 55 or about 25 to about 55 or about 35 to about 55 cell doublings after initiation of artificial chromosome generation, or they may appear after several cycles of growth under selective conditions and BrdU treatment.

As used herein, an artificial chromosome that is predominantly heterochromatic (*i.e.*, containing more heterochromatin than euchromatin, typically more than about 50%, more than about 70%, or more than about 90% heterochromatin) may be produced by introducing nucleic acid molecules into cells, such as, for example, animal or plant cells, and selecting cells that contain a predominantly heterochromatic artificial chromosome. Any nucleic acid may be introduced into cells in such methods of producing the artificial chromosomes. For example, the nucleic acid may contain a selectable marker and/or optionally a sequence that targets nucleic acid to the pericentric, heterochromatic region of a chromosome, such as in the short arm of acrocentric chromosomes and

-22-

nucleolar organizing regions. Targeting sequences include, but are not limited to, lambda phage DNA and rDNA for production of predominantly heterochromatic artificial chromosomes in eukaryotic cells.

After introducing the nucleic acid into cells, a cell containing a
5 predominantly heterochromatic artificial chromosome is selected. Such
cells may be identified using a variety of procedures. For example,
repeating units of heterochromatic DNA of these chromosomes may be
discerned by G-banding and/or fluorescence in situ hybridization (FISH)
techniques. Prior to such analyses, the cells to be analyzed may be
10 enriched with artificial chromosome-containing cells by sorting the cells
on the basis of the presence of a selectable marker, such as a reporter
protein, or by growing (culturing) the cells under selective conditions. It
is also possible, after introduction of nucleic acids into cells, to select
cells that have a multicentric, typically dicentric, chromosome, a formerly
15 multicentric (typically dicentric) chromosome and/or various
heterochromatic structures, such as a megachromosome and a sausage
chromosome, that contain a centromere and are predominantly
heterochromatic and to treat them such that desired artificial
chromosomes are produced. Cells containing a new chromosome are
20 selected. Conditions for generation of a desired structure include, but are
not limited to, further growth under selective conditions, introduction of
additional nucleic acid molecules and/or growth under selective conditions
and treatment with destabilizing agents, and other such methods (see
International PCT application No. WO 97/40183 and U.S. Patent Nos.
25 6,025,155 and 6,077,697).

As used herein, a "selectable marker" is a nucleic acid segment,
generally DNA, that allows one to select for or against a molecule or a
cell that contains it, often under particular conditions. These markers can
encode an activity, such as, but not limited to, production of RNA,

-23-

peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds and compositions. Examples of selectable markers include but are not limited to: (1) nucleic acid segments that encode products that provide resistance against otherwise
5 toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid segments that encode products that can be identified, such as phenotypic markers,
10 including β -galactosidase, red, blue and/or green fluorescent proteins (FPs), and cell surface proteins; (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described in Nos. 1-5 above (e.g., antisense
15 oligonucleotides or siRNA molecules for use in RNA interference); (7) nucleic acid segments that bind products that modify a substrate (e.g. restriction endonucleases); (8) nucleic acid segments that can be used to isolate a desired molecule (e.g. specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence that can be
20 otherwise non-functional, such as for PCR amplification of subpopulations of molecules; and/or (10) nucleic acid segments, which when absent, directly or indirectly confer sensitivity to particular compounds. Thus, for example, selectable markers include nucleic acids encoding fluorescent proteins, such as green fluorescent proteins, β -galactosidase and other
25 readily detectable proteins, such as chromogenic proteins or proteins capable of being bound by an antibody and FACs sorted. Selectable markers such as these, which are not required for cell survival and/or proliferation in the presence of a selection agent, are also referred to herein as reporter molecules. Other selectable markers, e.g., the

-24-

neomycin phosphotransferase gene, provide for isolation and identification of cells containing them by conferring properties on the cells that make them resistant to an agent, *e.g.*, a drug such as an antibiotic, that inhibits proliferation of cells that do not contain the marker.

- 5 As another example, interference of gene expression by double stranded RNA has been shown in *Caenorhabditis elegans*, plants, *Drosophila*, protozoans and mammals. This method is known as RNA interference (RNAi) and utilizes short, double-stranded RNA molecules (siRNAs). The siRNAs are generally composed of a 19-22bp double-
- 10 stranded RNA stem, a loop region and a 1-4 bp overhang on the 3' end. The reduction of gene expression has been accomplished by direct introduction of the siRNAs into the cell (Harborth J et al., 2001, J Cell Sci 114(pt 24):4557-65) as well as the introduction of DNA encoding and expressing the siRNA molecule. The encoded siRNA molecules are under
- 15 the regulation of an RNA polymerase III promoter (see, *e.g.*, Yu et al., 2002, Proc Natl Acad Sci USA 99(9):6047-52; Brummelkamp et al., 2002, Science 296(5567):550-3; Miyagishi et al., 2002, Nat Biotechnol 20(5):497-500; and the like). In certain embodiments, RNAi in mammalian cells may have advantages over other therapeutic methods.
- 20 For example, producing siRNA molecules that block viral genetic activities in infected cells may reduce the effects of the virus. Platform ACes provided herein encoding siRNA molecule(s) are an additional utilization of the platform ACes technology. The platform ACes could be engineered to encode one or more siRNA molecules to create gene "knockdowns". In
- 25 one embodiment, a platform ACes can engineered to encode both the siRNA molecule and a replacement gene. For example, a mouse model or cell culture system could be generated using a platform ACes that has a knockdown of the endogenous mouse gene, by siRNA, and the human gene homolog expressing in place of the mouse gene. The placement of

-25-

siRNA encoding sequences under the regulation of a regulatable or inducible promoter would allow one to temporally and/or spatially control the knockdown effect of the corresponding gene.

As used herein, a reporter gene includes any gene that expresses a detectable gene product, which may be RNA or protein. Generally reporter genes are readily detectable. Examples of reporter genes include, but are not limited to nucleic acid encoding a fluorescent protein, CAT (chloramphenicol acetyl transferase) (Alton *et al.* (1979) *Nature* 282: 864-869) luciferase, and other enzyme detection systems, such as beta-galactosidase; firefly luciferase (deWet *et al.* (1987) *Mol. Cell. Biol.* 7:725-737); bacterial luciferase (Engebrecht and Silverman (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81:4154-4158; Baldwin *et al.* (1984) *Biochemistry* 23:3663-3667); and alkaline phosphatase (Toh *et al.* (1989) *Eur. J. Biochem.* 182:231-238, Hall *et al.* (1983) *J. Mol. Appl. Gen.* 2:101).

As used herein, growth under selective conditions means growth of a cell under conditions that require expression of a selectable marker for survival.

As used herein, an agent that destabilizes a chromosome is any agent known by those skilled in the art to enhance amplification events, and/or mutations. Such agents, which include BrdU, are well known to those skilled in the art.

In order to generate an artificial chromosome containing a particular heterologous nucleic acid of interest, it is possible to include the nucleic acid in the nucleic acid that is being introduced into cells to initiate production of the artificial chromosome. Thus, for example, a nucleic acid can be introduced into a cell along with nucleic acid encoding a selectable marker and/or a nucleic acid that targets to a heterochromatic region of a chromosome. For introducing a heterologous nucleic acid into

-26-

the cell, it can be included in a fragment that includes a selectable marker or as part of a separate nucleic acid fragment and introduced into the cell with a selectable marker during the process of generating the artificial chromosomes. Alternatively, heterologous nucleic acid can be introduced
5 into an artificial chromosome at a later time after the initial generation of the artificial chromosome.

As used herein, the minichromosome refers to a chromosome derived from a multicentric, typically dicentric, chromosome that contains more euchromatic than heterochromatic DNA. For purposes herein, the
10 minichromosome contains a *de novo* centromere (e.g., a neocentromere). In some embodiments, for example, the minichromosome contains a centromere that replicates in animals, e.g., a mammalian centromere or in plants, e.g., a plant centromere.

As used herein, *in vitro* assembled artificial chromosomes or
15 synthetic chromosomes can be either more euchromatic than heterochromatic or more heterochromatic than euchromatic and are produced by joining essential components of a chromosome *in vitro*. These components include at least a centromere, a megareplicator, a telomere and optionally secondary origins of replication.

As used herein, *in vitro* assembled plant or animal artificial
20 chromosomes are produced by joining essential components (at least the centromere, telomere(s), megareplicator and optional secondary origins of replication) that function in plants or animals. In particular embodiments, the megareplicator contains sequences of rDNA, particularly plant or
25 animal rDNA.

As used herein, a plant is a eukaryotic organism that contains, in addition to a nucleus and mitochondria, chloroplasts capable of carrying out photosynthesis. A plant can be unicellular or multicellular and can contain multiple tissues and/or organs. Plants can reproduce sexually or

-27-

asexually and can be perennial or annual in growth. Plants can also be terrestrial or aquatic. The term "plant" includes a whole plant, plant cell, plant protoplast, plant calli, plant seed, plant organ, plant tissue, and other parts of a whole plant.

5 As used herein, stable maintenance of chromosomes occurs when at least about 85%, preferably 90%, more preferably 95%, of the cells retain the chromosome. Stability is measured in the presence of a selective agent. Preferably these chromosomes are also maintained in the absence of a selective agent. Stable chromosomes also retain their
10 structure during cell culturing, suffering no unintended intrachromosomal or interchromosomal rearrangements.

 As used herein, *de novo* with reference to a centromere, refers to generation of an excess centromere in a chromosome as a result of incorporation of a heterologous nucleic acid fragment using the methods
15 herein.

 As used herein, BrdU refers to 5-bromodeoxyuridine, which during replication is inserted in place of thymidine. BrdU is used as a mutagen; it also inhibits condensation of metaphase chromosomes during cell division.

20 As used herein, ribosomal RNA (rRNA) is the specialized RNA that forms part of the structure of a ribosome and participates in the synthesis of proteins. Ribosomal RNA is produced by transcription of genes which, in eukaryotic cells, are present in multiple copies. In human cells, the approximately 250 copies of rRNA genes (i.e., genes which encode rRNA)
25 per haploid genome are spread out in clusters on at least five different chromosomes (chromosomes 13, 14, 15, 21 and 22). In mouse cells, the presence of ribosomal DNA (rDNA, which is DNA containing sequences that encode rRNA) has been verified on at least 11 pairs out of 20 mouse chromosomes (chromosomes 5, 6, 7, 9, 11, 12, 15, 16, 17, 18, and 19)

-28-

(see e.g., Rowe *et al.* (1996) *Mamm. Genome* 7:886-889 and Johnson *et al.* (1993) *Mamm. Genome* 4:49-52). In *Arabidopsis thaliana* the presence of rDNA has been verified on chromosomes 2 and 4 (18S, 5.8S, and 25S rDNA) and on chromosomes 3,4, and 5 (5S rDNA)(see The
5 *Arabidopsis* Genome Initiative (2000) *Nature* 408:796-815). In eukaryotic cells, the multiple copies of the highly conserved rRNA genes are located in a tandemly arranged series of rDNA units, which are generally about 40-45 kb in length and contain a transcribed region and a nontranscribed region known as spacer (i.e., intergenic spacer) DNA
10 which can vary in length and sequence. In the human and mouse, these tandem arrays of rDNA units are located adjacent to the pericentric satellite DNA sequences (heterochromatin). The regions of these chromosomes in which the rDNA is located are referred to as nucleolar organizing regions (NOR) which loop into the nucleolus, the site of
15 ribosome production within the cell nucleus.

As used herein, a megachromosome refers to a chromosome that, except for introduced heterologous DNA, is substantially composed of heterochromatin. Megachromosomes are made up of an array of repeated amplicons that contain two inverted megareplicons bordered by
20 introduced heterologous DNA (see, *e.g.*, Figure 3 of U.S. Patent No. 6,077,697 for a schematic drawing of a megachromosome). For purposes herein, a megachromosome is about 50 to 400 Mb, generally about 250-400 Mb. Shorter variants are also referred to as truncated megachromosomes (about 90 to 120 or 150 Mb), dwarf
25 megachromosomes (~ 150-200 Mb), and a micro-megachromosome (~ 50-90 Mb, typically 50-60 Mb). For purposes herein, the term

-29-

megachromosome refers to the overall repeated structure based on an array of repeated chromosomal segments (amplicons) that contain two inverted megareplicons bordered by any inserted heterologous DNA. The size will be specified.

5 As used herein, gene therapy involves the transfer or insertion of nucleic acid molecules into certain cells, which are also referred to as target cells, to produce specific products that are involved in preventing, curing, correcting, controlling or modulating diseases, disorders and deleterious conditions. The nucleic acid is introduced into the selected
10 target cells in a manner such that the nucleic acid is expressed and a product encoded thereby is produced. Alternatively, the nucleic acid may in some manner mediate expression of DNA that encodes a therapeutic product. This product may be a therapeutic compound, which is produced in therapeutically effective amounts or at a therapeutically
15 useful time. It may also encode a product, such as a peptide or RNA, that in some manner mediates, directly or indirectly, expression of a therapeutic product. Expression of the nucleic acid by the target cells within an organism afflicted with a disease or disorder thereby provides for modulation of the disease or disorder. The nucleic acid encoding the
20 therapeutic product may be modified prior to introduction into the cells of the afflicted host in order to enhance or otherwise alter the product or expression thereof.

For use in gene therapy, cells can be transfected *in vitro*, followed by introduction of the transfected cells into an organism. This is often
25 referred to as *ex vivo* gene therapy. Alternatively, the cells can be transfected directly *in vivo* within an organism.

As used herein, therapeutic agents include, but are not limited to, growth factors, antibodies, cytokines, such as tumor necrosis factors and interleukins, and cytotoxic agents and other agents disclosed herein and

-30-

known to those of skill in the art. Such agents include, but are not limited to, tumor necrosis factor, α -interferon, β -interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines,
5 interleukin-1 (IL-1), interleukin-2 (IL-2), interleukin-6 (IL-6), granulocyte macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), erythropoietin (EPO), pro-coagulants such as tissue factor and tissue factor variants, pro-apoptotic agents such as FAS-ligand, fibroblast growth factors (FGF), nerve growth factor and other
10 growth factors.

As used herein, a therapeutically effective product is a product that is encoded by heterologous DNA that, upon introduction of the DNA into a host, a product is expressed that effectively ameliorates or eliminates the symptoms, manifestations of an inherited or acquired disease or that
15 cures the disease.

As used herein, transgenic plants and animals refer to plants and animals in which heterologous or foreign nucleic acid is expressed or in which the expression of a gene naturally present in the plant or animal has been altered by virtue of introduction of heterologous or foreign
20 nucleic acid.

As used herein, IRES (internal ribosome entry site; see, *e.g.*, SEQ ID No. 27 and nucleotides 2736-3308 SEQ ID No. 28) refers to a region of a nucleic acid molecule, such as an mRNA molecule, that allows internal ribosome entry sufficient to initiate translation, which initiation
25 can be detected in an assay for cap-independent translation (see, *e.g.*, U.S. Patent No. 6,171,821). The presence of an IRES within an mRNA molecule allows cap-independent translation of a linked protein-encoding sequence that otherwise would not be translated.

-31-

Internal ribosome entry site (IRES) elements were first identified in picornaviruses, which elements are considered the paradigm for cap-independent translation. The 5' UTRs of all picornaviruses are long and mediate translational initiation by directly recruiting and binding
5 ribosomes, thereby circumventing the initial cap-binding step. IRES elements are frequently found in viral mRNA, they are rare in non-viral mRNA. Among non-viral mRNA molecules that contain functional IRES elements in their respective 5' UTRs are those encoding immunoglobulin heavy chain binding protein (BiP) (Macejak *et al.* (1991) *Nature*
10 353:90-94); *Drosophila* Antennapedia (Oh *et al.* (1992) *Genes Dev.*, 6:1643-1653); D. Ultrabithorax (Ye *et al.* (1997) *Mol. Cell Biol.* 17:1714-21); fibroblast growth factor 2 (Vagner *et al.* (1995) *Mol. Cell Biol.* 15:35-44); initiation factor eIF4G (Gan *et al.* (1998) *J. Biol. Chem.* 273:5006-5012); proto-oncogene c-myc (Nanbru *et al.* (1995) *J. Biol.*
15 *Chem.* 272:32061-32066; Stoneley (1998) *Oncogene* 16:423-428); IRES_H; from the 5'UTR of NRF1 gene (Oumard *et al.* (2000) *Mol. and Cell Biol.*, 20(8):2755-2759); and vascular endothelial growth factor (VEGF) (Stein *et al.* (1998) *Mol. Cell Biol.* 18:3112-9).

As used herein, a promoter, with respect to a region of DNA, refers
20 to a sequence of DNA that contains a sequence of bases that signals RNA polymerase to associate with the DNA and initiate transcription of RNA (such as pol II for mRNA) from a template strand of the DNA. A promoter thus generally regulates transcription of DNA into mRNA. A particular promoter provided herein is the Ferritin heavy chain promoter (excluding
25 the Iron Response Element, located in the 5'UTR), which was joined to the 37bp Fer-1 enhancer element. This promoter is set forth as SEQ ID NO:128. The endogenous Fer-1 enhancer element is located upstream of the Fer-1 promoter (e.g., a Fer-1 oligo was cloned proximal to the core promoter).

-32-

As used herein, isolated, substantially pure nucleic acid, such as, for example, DNA, refers to nucleic acid fragments purified according to standard techniques employed by those skilled in the art, such as that found in Sambrook *et al.* ((2001) Molecular Cloning: A Laboratory
5 Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 3rd edition).

As used herein, expression refers to the transcription and/or translation of nucleic acid. For example, expression can be the transcription of a gene that may be transcribed into an RNA molecule,
10 such as a messenger RNA (mRNA) molecule. Expression may further include translation of an RNA molecule and translated into peptides, polypeptides, or proteins. If the nucleic acid is derived from genomic DNA, expression may, if an appropriate eukaryotic host cell or organism is selected, include splicing of the mRNA. With respect to an antisense
15 construct, expression may refer to the transcription of the antisense DNA.

As used herein, vector or plasmid refers to discrete elements that are used to introduce heterologous nucleic acids into cells for either expression of the heterologous nucleic acid or for replication of the heterologous nucleic acid. Selection and use of such vectors and
20 plasmids are well within the level of skill of the art.

As used herein, transformation/transfection refers to the process by which nucleic acid is introduced into cells. The terms transfection and transformation refer to the taking up of exogenous nucleic acid, e.g., an expression vector, by a host cell whether or not any coding sequences
25 are in fact expressed. Numerous methods of transfection are known to the ordinarily skilled artisan, for example, by *Agrobacterium*-mediated transformation, protoplast transformation (including polyethylene glycol (PEG)-mediated transformation, electroporation, protoplast fusion, and microcell fusion), lipid-mediated delivery, liposomes, electroporation,

-33-

sonoporation, microinjection, particle bombardment and silicon carbide whisker-mediated transformation and combinations thereof (see, *e.g.*, Paszkowski *et al.* (1984) *EMBO J.* 3:2717-2722; Potrykus *et al.* (1985) *Mol. Gen. Genet.* 199:169-177; Reich *et al.* (1986) *Biotechnology* 4:1001-1004; Klein *et al.* (1987) *Nature* 327:70-73; U.S. Patent No. 6,143,949; Paszkowski *et al.* (1989) in *Cell Culture and Somatic Cell Genetics of Plants, Vol. 6*, Molecular Biology of Plant Nuclear Genes, eds. Schell, J and Vasil, L.K. Academic Publishers, San Diego, California, p. 52-68; and Frame *et al.* (1994) *Plant J.* 6:941-948), direct uptake using calcium phosphate (CaPO₄; see, *e.g.*, Wigler *et al.* (1979) *Proc. Natl. Acad. Sci. U.S.A.* 76:1373-1376), polyethylene glycol (PEG)-mediated DNA uptake, lipofection (see, *e.g.*, Strauss (1996) *Meth. Mol. Biol.* 54:307-327), microcell fusion (see, EXAMPLES, see, also Lambert (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:5907-5911; U.S. Patent No. 5,396,767, Sawford *et al.* (1987) *Somatic Cell Mol. Genet.* 13:279-284; Dhar *et al.* (1984) *Somatic Cell Mol. Genet.* 10:547-559; and McNeill-Killary *et al.* (1995) *Meth. Enzymol.* 254:133-152), lipid-mediated carrier systems (see, *e.g.*, Teifel *et al.* (1995) *Biotechniques* 19:79-80; Albrecht *et al.* (1996) *Ann. Hematol.* 72:73-79; Holmen *et al.* (1995) *In Vitro Cell Dev. Biol. Anim.* 31:347-351; Remy *et al.* (1994) *Bioconjug. Chem.* 5:647-654; Le Bolch *et al.* (1995) *Tetrahedron Lett.* 36:6681-6684; Loeffler *et al.* (1993) *Meth. Enzymol.* 217:599-618) or other suitable method. Methods for delivery of ACes are described in copending U.S. application Serial No. 09/815,979. Successful transfection is generally recognized by detection of the presence of the heterologous nucleic acid within the transfected cell, such as, for example, any visualization of the heterologous nucleic acid or any indication of the operation of a vector within the host cell.

-34-

As used herein, "delivery," which is used interchangeably with "transfection," refers to the process by which exogenous nucleic acid molecules are transferred into a cell such that they are located inside the cell. Delivery of nucleic acids is a distinct process from expression of
5 nucleic acids.

As used herein, injected refers to the microinjection, such as by use of a small syringe, needle, or pipette, for injection of nucleic acid into a cell.

As used herein, substantially homologous DNA refers to DNA that
10 includes a sequence of nucleotides that is sufficiently similar to another such sequence to form stable hybrids, with each other or a reference sequence, under specified conditions.

It is well known to those of skill in this art that nucleic acid fragments with different sequences may, under the same conditions,
15 hybridize detectably to the same "target" nucleic acid. Two nucleic acid fragments hybridize detectably, under stringent conditions over a sufficiently long hybridization period, because one fragment contains a segment of at least about 10, 14 or 16 or more nucleotides in a sequence that is complementary (or nearly complementary) to a substantially
20 contiguous sequence of at least one segment in the other nucleic acid fragment. If the time during which hybridization is allowed to occur is held constant, at a value during which, under preselected stringency conditions, two nucleic acid fragments with complementary base-pairing segments hybridize detectably to each other, departures from exact
25 complementarity can be introduced into the base-pairing segments, and base-pairing will nonetheless occur to an extent sufficient to make hybridization detectable. As the departure from complementarity between the base-pairing segments of two nucleic acids becomes larger, and as

-35-

conditions of the hybridization become more stringent, the probability decreases that the two segments will hybridize detectably to each other.

Two single-stranded nucleic acid segments have "substantially the same sequence", if (a) both form a base-paired duplex with the same
5 segment, and (b) the melting temperatures of the two duplexes in a solution of 0.5 X SSPE differ by less than 10°C. If the segments being compared have the same number of bases, then to have "substantially the same sequence", they will typically differ in their sequences at fewer than 1 base in 10. Methods for determining melting temperatures of
10 nucleic acid duplexes are well known (see, *e.g.*, Meinkoth *et al.* (1984) *Anal. Biochem.* 138:267-284 and references cited therein).

As used herein, a nucleic acid probe is a DNA or RNA fragment that includes a sufficient number of nucleotides to specifically hybridize to DNA or RNA that includes complementary or substantially complementary
15 sequences of nucleotides. A probe may contain any number of nucleotides, from as few as about 10 and as many as hundreds of thousands of nucleotides. The conditions and protocols for such hybridization reactions are well known to those of skill in the art as are the effects of probe size, temperature, degree of mismatch, salt
20 concentration and other parameters on the hybridization reaction. For example, the lower the temperature and higher the salt concentration at which the hybridization reaction is carried out, the greater the degree of mismatch that may be present in the hybrid molecules.

To be used as a hybridization probe, the nucleic acid is generally
25 rendered detectable by labeling it with a detectable moiety or label, such as ³²P, ³H and ¹⁴C, or by other means, including chemical labeling, such as by nick-translation in the presence of deoxyuridylate biotinylated at the 5'-position of the uracil moiety. The resulting probe includes the biotinylated uridylate in place of thymidylate residues and can be detected

-36-

(via the biotin moieties) by any of a number of commercially available detection systems based on binding of streptavidin to the biotin. Such commercially available detection systems can be obtained, for example, from Enzo Biochemicals, Inc. (New York, NY). Any other label known to those of skill in the art, including non-radioactive labels, may be used as long as it renders the probes sufficiently detectable, which is a function of the sensitivity of the assay, the time available (for culturing cells, extracting DNA, and hybridization assays), the quantity of DNA or RNA available as a source of the probe, the particular label and the means used to detect the label.

Once sequences with a sufficiently high degree of homology to the probe are identified, they can readily be isolated by standard techniques (see, *e.g.*, Sambrook *et al.* (2001) *Molecular Cloning: A Laboratory Manual*, 3rd Edition, Cold Spring Harbor Laboratory Press).

As used herein, conditions under which DNA molecules form stable hybrids are considered substantially homologous, and a DNA or nucleic acid homolog refers to a nucleic acid that includes a preselected conserved nucleotide sequence, such as a sequence encoding a polypeptide. By the term "substantially homologous" is meant having at least 75%, preferably 80%, preferably at least 90%, most preferably at least 95% homology therewith or a less percentage of homology or identity and conserved biological activity or function.

The terms "homology" and "identity" are often used interchangeably. In this regard, percent homology or identity may be determined, for example, by comparing sequence information using a GAP computer program. The GAP program utilizes the alignment method of Needleman and Wunsch (*J. Mol. Biol.* 48:443 (1970)), as revised by Smith and Waterman (*Adv. Appl. Math.* 2:482 (1981)). Briefly, the GAP program defines similarity as the number of aligned symbols (*i.e.*,

-37-

nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences. The preferred default parameters for the GAP program may include: (1) a unary comparison matrix (containing a value of 1 for identities and 0 for non-identities) and
5 the weighted comparison matrix of Gribskov and Burgess, *Nucl. Acids Res.* 14:6745 (1986), as described by Schwartz and Dayhoff, eds., *ATLAS OF PROTEIN SEQUENCE AND STRUCTURE*, National Biomedical Research Foundation, pp. 353-358 (1979); (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3)
10 no penalty for end gaps.

By sequence identity, the number of conserved amino acids are determined by standard alignment algorithms programs, and are used with default gap penalties established by each supplier. Substantially homologous nucleic acid molecules would hybridize typically at moderate
15 stringency or at high stringency all along the length of the nucleic acid of interest. Preferably the two molecules will hybridize under conditions of high stringency. Also contemplated are nucleic acid molecules that contain degenerate codons in place of codons in the hybridizing nucleic acid molecule.

20 Whether any two nucleic acid molecules have nucleotide sequences that are at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% "identical" can be determined using known computer algorithms such as the "FAST A" program, using for example, the default parameters as in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 (1988).

25 Alternatively the BLAST function of the National Center for Biotechnology Information database may be used to determine relative sequence identity.

In general, sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be

-38-

calculated using published techniques. (See, *e.g.*: *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H. & Lipton, D., *SIAM J Applied Math* 48:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H. & Lipton, D., *SIAM J Applied Math* 48:1073 (1988). Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., *et al.*, *Nucleic Acids Research* 12(II):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., *et al.*, *J Molec Biol* 215:403 (1990)).

Therefore, as used herein, the term "identity" represents a comparison between a test and a reference polypeptide or polynucleotide. For example, a test polypeptide may be defined as any polypeptide that is 90% or more identical to a reference polypeptide.

As used herein, the term at least "90% identical to" refers to percent identities from 90 to 99.99 relative to the reference polypeptides. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of

-39-

100 amino acids are compared. No more than 10% (i.e., 10 out of 100) amino acids in the test polypeptide differs from that of the reference polypeptides. Similar comparisons may be made between a test and reference polynucleotides. Such differences may be represented as point mutations randomly distributed over the entire length of an amino acid sequence or they may be clustered in one or more locations of varying length up to the maximum allowable, e.g. 10/100 amino acid difference (approximately 90% identity). Differences are defined as nucleic acid or amino acid substitutions, or deletions.

10 As used herein: stringency of hybridization in determining percentage mismatch encompass the following conditions or equivalent conditions thereto:

- 1) high stringency: 0.1 x SSPE or SSC, 0.1% SDS, 65°C
- 2) medium stringency: 0.2 x SSPE or SSC, 0.1% SDS, 50°C
- 15 3) low stringency: 1.0 x SSPE or SSC, 0.1% SDS, 50°C

or any combination of salt and temperature and other reagents that result in selection of the same degree of mismatch or matching. Equivalent conditions refer to conditions that select for substantially the same percentage of mismatch in the resulting hybrids. Additions of ingredients, such as formamide, Ficoll, and Denhardt's solution affect parameters such as the temperature under which the hybridization should be conducted and the rate of the reaction. Thus, hybridization in 5 X SSC, in 20% formamide at 42° C is substantially the same as the conditions recited above hybridization under conditions of low stringency. The recipes for 25 SSPE, SSC and Denhardt's and the preparation of deionized formamide are described, for example, in Sambrook *et al.* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Chapter 8; see, Sambrook *et al.*, vol. 3, p. B.13, see, also, numerous catalogs that describe commonly used laboratory solutions. It is understood that

-40-

equivalent stringencies may be achieved using alternative buffers, salts and temperatures. As used herein, all assays and procedures, such as hybridization reactions and antibody-antigen reactions, unless otherwise specified, are conducted under conditions recognized by those of skill in the art as standard conditions.

As used herein, conservative amino acid substitutions, such as those set forth in Table 1, are those that do not eliminate biological activity. Suitable conservative substitutions of amino acids are known to those of skill in this art and may be made generally without altering the biological activity of the resulting molecule. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (see, *e.g.*, Watson *et al. Molecular Biology of the Gene*, 4th Edition, 1987, The Bejacmin/Cummings Pub. co., p.224). Conservative amino acid substitutions are made, for example, in accordance with those set forth in TABLE 1 as follows:

TABLE 1

	Original residue	Conservative substitution
	Ala (A)	Gly; Ser, Abu
20	Arg (R)	Lys, orn
	Asn (N)	Gln; His
	Cys (C)	Ser
	Gln (Q)	Asn
	Glu (E)	Asp
25	Gly (G)	Ala; Pro
	His (H)	Asn; Gln
	Ile (I)	Leu; Val; Met; Nle; Nva
	Leu (L)	Ile; Val; Met; Nle; Nva
	Lys (K)	Arg; Gln; Glu
30	Met (M)	Leu; Tyr; Ile; NLe Val
	Ornithine	Lys; Arg
	Phe (F)	Met; Leu; Tyr
	Ser (S)	Thr
	Thr (T)	Ser
35	Trp (W)	Tyr
	Tyr (Y)	Trp; Phe
	Val (V)	Ile; Leu; Met; Nle; Nva

-41-

Other substitutions are also permissible and may be determined empirically or in accord with known conservative substitutions.

As used herein, the amino acids, which occur in the various amino acid sequences appearing herein, are identified according to their well-known, three-letter or one-letter abbreviations. The nucleotides, which
5 occur in the various DNA fragments, are designated with the standard single-letter designations used routinely in the art.

As used herein, a splice variant refers to a variant produced by differential processing of a primary transcript of genomic DNA that results
10 in more than one type of mRNA.

As used herein, a probe or primer based on a nucleotide sequence includes at least 10, 14, 16, 30 or 100 contiguous nucleotides from the reference nucleic acid molecule.

As used herein, recombinant production by using recombinant DNA
15 methods refers to the use of the well known methods of molecular biology for expressing proteins encoded by cloned DNA.

As used herein, biological activity refers to the *in vivo* activities of a compound or physiological responses that result upon *in vivo* administration of a compound, composition or other mixture. Biological
20 activity, thus, encompasses therapeutic effects and pharmaceutical activity of such compounds, compositions and mixtures. Biological activities may be observed in *in vitro* systems designed to test or use such activities. Thus, for purposes herein the biological activity of a luciferase is its oxygenase activity whereby, upon oxidation of a
25 substrate, light is produced.

-42-

The terms substantially identical or similar varies with the context as understood by those skilled in the relevant art and generally means at least 40, 60, 80, 90, 95 or 98%.

As used herein, substantially identical to a product means
5 sufficiently similar so that the property is sufficiently unchanged so that the substantially identical product can be used in place of the product.

As used herein, substantially pure means sufficiently homogeneous to appear free of readily detectable impurities as determined by standard methods of analysis, such as thin layer chromatography (TLC), gel
10 electrophoresis and high performance liquid chromatography (HPLC), used by those of skill in the art to assess such purity, or sufficiently pure such that further purification would not detectably alter the physical and chemical properties, such as enzymatic and biological activities, of the substance. Methods for purification of the compounds to produce
15 substantially chemically pure compounds are known to those of skill in the art. A substantially chemically pure compound may, however, be a mixture of stereoisomers or isomers. In such instances, further purification might increase the specific activity of the compound.

As used herein, vector (or plasmid) refers to discrete elements that
20 are used to introduce heterologous DNA into cells for either expression or replication thereof. The vectors typically remain episomal, but may be designed to effect integration of a gene or portion thereof into a chromosome of the genome. Also contemplated are vectors that are artificial chromosomes, such as yeast artificial chromosomes and
25 mammalian artificial chromosomes. Selection and use of such vehicles are well known to those of skill in the art. An expression vector includes vectors capable of expressing DNA that is operatively linked with regulatory sequences, such as promoter regions, that are capable of effecting expression of such DNA fragments. Thus, an expression vector

-43-

refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector that, upon introduction into an appropriate host cell, results in expression of the cloned DNA.

Appropriate expression vectors are well known to those of skill in the art
5 and include those that are replicable in eukaryotic cells and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome.

As used herein, protein-binding-sequence refers to a protein or peptide sequence that is capable of specific binding to other protein or
10 peptide sequences generally, to a set of protein or peptide sequences or to a particular protein or peptide sequence.

As used herein, a composition refers to any mixture of two or more ingredients. It may be a solution, a suspension, liquid, powder, a paste, aqueous, non-aqueous or any combination thereof.

15 As used herein, a combination refers to any association between two or more items.

As used herein, fluid refers to any composition that can flow. Fluids thus encompass compositions that are in the form of semi-solids, pastes, solutions, aqueous mixtures, gels, lotions, creams and other such
20 compositions.

As used herein, a cellular extract refers to a preparation or fraction that is made from a lysed or disrupted cell.

As used herein, the term "subject" refers to animals, plants, insects, and birds and other phyla, genera and species into which nucleic
25 acid molecules may be introduced. Included are higher organisms, such as mammals, fish, insects and birds, including humans, primates, cattle, pigs, rabbits, goats, sheep, mice, rats, guinea pigs, hamsters, cats, dogs, horses, chicken and others.

-44-

As used herein, flow cytometry refers to processes that use a laser based instrument capable of analyzing and sorting out cells and or chromosomes based on size and fluorescence.

As used herein, the abbreviations for any protective groups, amino
5 acids and other compounds, are, unless indicated otherwise, in accord with their common usage, recognized abbreviations, or the IUPAC-IUB Commission on Biochemical Nomenclature (see, (1972) *Biochem.* 11:942-944).

B. Recombination systems

10 Site-specific recombination systems typically contain three elements: a pair of DNA sequences (the site-specific recombination sequences) and a specific enzyme (the site-specific recombinase). The site-specific recombinase catalyzes a recombination reaction between two site-specific recombination sequences.

15 A number of different site-specific recombinase systems are available and/or known to those of skill in the art, including, but not limited to: the Cre//lox recombination system using CRE recombinase (see, *e.g.*, SEQ ID Nos. 58 and 59) from the Escherichia coli phage P1 (see, *e.g.*, Sauer (1993) *Methods in Enzymology* 225:890-900; Sauer *et al.* 20 (1990) *The New Biologist* 2:441-449), Sauer (1994) *Current Opinion in Biotechnology* 5:521-527; Odell *et al.* (1990) *Mol Gen Genet.* 223:369-378; Lasko *et al.* (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:6232-6236; U.S. Patent No. 5,658,772), the FLP/FRT system of yeast using the FLP recombinase (see, SEQ ID Nos. 60 and 61) from the 2 μ episome of 25 *Saccharomyces cerevisiae* (Cox (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80:4223; Falco *et al.* (1982) *Cell* 29:573-584; Golic *et al.* (1989) *Cell* 59:499-509; U.S. Patent No. 5,744,336), the resolvases, including Gin recombinase of phage Mu (Maeser *et al.* (1991) *Mol Gen Genet.* 230:170-176; Klippel, A. *et al.* (1993) *EMBO J.* 12:1047-1057; see, *e.g.*,

-45-

SEQ ID Nos. 64-67), Cin, Hin, $\alpha\delta$ Tn3; the Pin recombinase of *E. coli* (see, e.g., SEQ ID Nos. 68 and 69; Enomoto *et al.* (1983) *J. Bacteriol.* 6:663-668), the R/RS system of the pSR1 plasmid of *Zygosaccharomyces rouxii* (Araki *et al.* (1992) *J. Mol. Biol.* 225:25-37; Matsuzaki *et al.* (1990) *J. Bacteriol.* 172: 610-618) and site-specific recombinases from *Kluyveromyces drosophilarius* (Chen *et al.* (1986) *Nucleic Acids Res.* 314:4471-4481) and *Kluyveromyces waltii* (Chen *et al.* (1992) *J. Gen. Microbiol.* 138:337-345). Other systems are known to those of skill in the art (Stark *et al.* *Trends Genet.* 8:432-439; Utatsu *et al.* (1987) *J. Bacteriol.* 169:5537-5545; see, also, U.S. Patent No. 6,171,861).

Members of the highly related family of site-specific recombinases, the resolvase family, such as $\gamma\delta$, Tn3 resolvase, Hin, Gin, and Cin are also available. Members of this family of recombinases are typically constrained to intramolecular reactions (e.g., inversions and excisions) and can require host-encoded factors. Mutants have been isolated that relieve some of the requirements for host factors (Maeser *et al.* (1991) *Mol. Gen. Genet.* 230:170-176), as well as some of the constraints of intramolecular recombination (see, U.S. Patent No. 6,171,861).

The bacteriophage P1 Cre/lox and the yeast FLP/FRT systems are particularly useful systems for site-specific integration, inversion or excision of heterologous nucleic acid into, and out of, chromosomes, particularly ACes as provided herein. In these systems a recombinase (Cre or FLP) interacts specifically with its respective site-specific recombination sequence (lox or FRT, respectively) to invert or excise the intervening sequences. The sequence for each of these two systems is relatively short (34 bp for lox and 47 bp for FRT).

The FLP/FRT recombinase system has been demonstrated to function efficiently in plant cells (U.S. Patent No. 5,744,386), and, thus, can be used for producing plant artificial chromosome platforms. In

-46-

general, short incomplete FRT sites leads to higher accumulation of excision products than the complete full-length FRT sites. The system catalyzes intra- and intermolecular reactions, and, thus, can be used for DNA excision and integration reactions. The recombination reaction is
5 reversible and this reversibility can compromise the efficiency of the reaction in each direction. Altering the structure of the site-specific recombination sequences is one approach to remedying this situation. The site-specific recombination sequence can be mutated in a manner that the product of the recombination reaction is no longer recognized as
10 a substrate for the reverse reaction, thereby stabilizing the integration or excision event.

In the Cre-lox system, discovered in bacteriophage P1, recombination between loxP sites occurs in the presence of the Cre recombinase (see, *e.g.*, U.S. Patent No. 5,658,772). This system can be
15 used to insert, invert or excise nucleic acid located between two lox sites. Cre can be expressed from a vector. Since the lox site is an asymmetrical nucleotide sequence, lox sites on the same DNA molecule can have the same or opposite orientation with respect to each other. Recombination between lox sites in the same orientation results in a deletion of the DNA
20 segment located between the two lox sites and a connection between the resulting ends of the original DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule
25 result in an inversion of the nucleotide sequence of the DNA segment located between the two lox sites. In addition, reciprocal exchange of DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the product of the Cre coding region.

-47-

Any site-specific recombinase system known to those of skill in the art is contemplated for use herein. It is contemplated that one or a plurality of sites that direct the recombination by the recombinase are introduced into an artificial chromosome to produce platform *ACes*. The
5 resulting platform *ACes* are introduced into cells with nucleic acid encoding the cognate recombinase, typically on a vector, and nucleic acid encoding heterologous nucleic acid of interest linked to the appropriate recombination site for insertion into the platform *ACes*. The recombinase-encoding-nucleic acid may be introduced into the cells on the same
10 vector, or a different vector, encoding the heterologous nucleic acid.

An *E. coli* phage lambda integrase system for *ACes* platform engineering and for artificial chromosome engineering is provided (Lorbach *et al.* (2000) *J. Mol. Biol* 296:1175-1181). The phage lambda integrase (Landy, A. (1989) *Annu. Rev. Biochem.* 58:913-94) is adapted herein and
15 the cognate *att* sites are provided. Chromosomes, including *ACes*, engineered to contain one or a plurality of *att* sites are provided, as are vectors encoding a mutant integrase that functions in the absence other factors. Methods using the modified chromosomes and vectors for introduction of heterologous nucleic acid are also provided.

20 For purposes herein, one or more of the sites (e.g., a single site or a pair of sites) required for recombination are introduced into an artificial chromosome, such as an *ACes* chromosome. The enzyme for catalyzing site-directed recombination is introduced with the DNA of interest, or separately, or is engineered onto the artificial chromosome under the
25 control of a regulatable promoter.

As described herein, artificial chromosome platforms containing one or multiple recombination sites are provided. The methods and resulting products are exemplified with the lambda phage Att/Int system, but

-48-

similar methods may be used for production of *ACes* platforms with other recombination systems.

The Att/Int system and vectors provided herein are not only intended for engineering *ACes* platforms, but may be used to engineer an Att/Int system into any chromosome. Introduction of att sites into a chromosome will permit engineering of natural chromosomes, such as by permitting targeted integration genes or regulatory regions, and by controlled excision of selected regions. For example, genes encoding a particular trait may be added to a chromosome, such as plant chromosome engineered to contain one or plurality of *att* sites. Such chromosomes may be used for screening DNA to identify genes. Large pieces of DNA can be introduced into cells and the cells screened phenotypically to select those having the desired trait.

C. Platforms

Provided herein are platform artificial chromosomes (platform *ACes*) containing single or multiple site-specific recombination sites. Chromosome-based platform technology permits efficient and tractable engineering and subsequent expression of multiple gene targets. Methods are provided that use DNA vectors and fragments to create platform artificial chromosomes, including animal, particularly mammalian, artificial chromosomes, and plant artificial chromosomes. The artificial chromosomes contain either single or multiple sequence-specific recombination sites suitable for the placement of target gene expression vectors onto the platform chromosome. The engineered chromosome-based platform *ACes* technology is applicable for methods, including cellular and transgenic protein production, transgenic plant and animal production and gene therapy. The platform *ACes* are also useful for producing a library of *ACes* comprising random portions of a given genome (e.g., a mammalian, plant or prokaryotic genome) for genomic

-49-

screening; as well as a library of cells comprising different and/or mutually exclusive *ACes* therein.

Exemplary of artificial chromosome platforms are those based on *ACes*. *ACes* artificial chromosomes are non-viral, self-replicating nucleic acid molecules that function as a natural chromosome, having all the elements required for normal chromosomal replication and maintenance within the cell nucleus. *ACes* artificial chromosomes do not rely on integration into the genome of the cell to be effective, and they are not limited by DNA carrying capacity and as such the therapeutic gene(s) of interest, including regulatory sequences, can be engineered into the *ACes*. In addition, *ACes* are stable *in vitro* and *in vivo* and can provide predictable long-term gene expression. Once engineered and delivered to the appropriate cell or embryo, *ACes* work independently alongside host chromosomes, for *ACes* that are predominantly heterochromatin producing only the products (proteins) from the genes it carries. As provided herein *ACes* are modified by introduction of recombination site(s) to provide a platform for ready introduction of heterologous nucleic acid. The *ACes* platforms can be used for production of transgenic animals and plants; as vectors for genetic therapy; for use as protein production systems; for animal models to identify and target new therapeutics; in cell culture for the development and production of therapeutic proteins; and for a variety of other applications.

1. Generation of artificial chromosomes

Artificial chromosomes may be generated by any method known to those of skill in the art. Of particular interest herein are the *ACes* artificial chromosomes, which contain a repeated unit. Methods for production of *ACes* are described in detail in U.S. Patent Nos. 6,025,155 and 6,077,697, which, as with all patents, applications, publications and other disclosure, are incorporated herein in their entirety.

-50-

Generation of *de novo* ACes.

ACes can be generated by cotransfecting exogenous DNA—such as a mammary tissue specific DNA cassette including the gene sequences for a therapeutic protein, with a rDNA fragment and a drug resistance
5 marker gene into the desired eukaryotic cell, such as plant or animal cells, such as murine cells *in vitro*. DNA with a selectable or detectable marker is introduced, and can be allowed to integrate randomly into pericentric heterochromatin or can be targeted to pericentric heterochromatin, such as that in rDNA gene arrays that reside on acrocentric chromosomes,
10 such as the short arms of acrocentric chromosomes. This integration event activates the “megareplicator” sequence and amplifies the pericentric heterochromatin and the exogenous DNA, and duplicates a centromere. Ensuing breakage of this “dicentric” chromosome can result in the production of daughter cells that contain the substantially-original
15 chromosome and the new artificial chromosome. The resulting ACes contain all the essential elements needed for stability and replication in dividing cells—centromere, origins of replications, and telomeres. ACes have been produced that express marker genes (lacZ, green fluorescent protein, neomycin-resistance, puromycin-resistance, hygromycin-
20 resistance) and genes of interest. Isolated ACes, for example, have been successfully transferred intact to rodent, human, and bovine cells by electroporation, sonoporation, microinjection, and transfection with lipids and dendrimers.

To render the creation of ACes with desired genes more tractable
25 and efficient, “platform” ACes (platform-ACes) can be produced that contain defined DNA sequences for enzyme-mediated homologous DNA recombination, such as by Cre or FLP recombinases (Bouhassira *et al.* (1996) *Blood* 88(supplement 1):190a; Bouhassira *et al.* (1997) *Blood*, 90:3332-3344; Siebler *et al.* (1997) *Biochemistry*: 36:1740-1747;

-51-

Siebler *et al.* (1998) *Biochemistry* 37: 6229-6234; and Bethke *et al.* (1997) *Nucl. Acids Res.* 25:2828-2834), and as exemplified herein the lambda phage integrase. A *lox* site contains two 13 bp inverted repeats to which Cre-recombinase binds and an intervening 8 bp core region.

- 5 Only pairs of sites having identity in the central 6 bp of the core region are proficient for recombination; sites having non-identical core sequences (heterospecific *lox* sites) do not efficiently recombine with each other (Hoess *et al.* (1986) *Nucleic Acids Res.* 14:2287-2300).

10 **Generating acrocentric chromosomes for plant artificial chromosome formation.**

In human and mouse cells *de novo* formation of a satellite DNA based artificial chromosome (SATAC, also referred to as *ACes*) can occur in an acrocentric chromosome where the short arm contains only pericentric heterochromatin, the rDNA array, and telomere sequences.

- 15 Plant species may not have any acrocentric chromosomes with the same physical structure described, but "megareplicator" DNA sequences reside in the plant rDNA arrays, also known as the nucleolar organizing regions (NOR). A structure like those seen in acrocentric mammalian chromosomes can be generated using site-specific recombination between
20 appropriate arms of plant chromosomes.

Approach

- Qin *et al.* ((1994) *Proc. Natl. Acad. Sci. U.S.A.* 91:1706-1710, 1994) describes crossing two *Nicotiana tabacum* transgenic plants. One plant contains a construct encoding a promoterless hygromycin-resistance gene preceded by a *lox* site (*lox-hpt*), the other plant carries a construct
25 containing a cauliflower mosaic virus 35S promoter linked to a *lox* sequence and the *cre* DNA recombinase coding region (35S-*lox-cre*). The constructs were introduced separately by infecting leaf explants with *agrobacterium tumefaciens* which carries the kanamycin-resistance gene

-52-

(Kan^R). The resultant Kan^R transgenic plants were crossed. Plants that carried the appropriate DNA recombination event were identified by hygromycin-resistance.

5 **Modification of the above for generation of ACes**

The Kan^R cultivars are initially screened, such as by FISH, to identify two sets of candidate transgenic plants. One set has one construct integrated in regions adjacent to the pericentric heterochromatin on the short arm of any chromosome. The second set of candidate plants
10 has the other construct integrated in the NOR region of appropriate chromosomes. To obtain reciprocal translocation both sites must be in the same orientation. Therefore a series of crosses are required, Kan^R plants generated, and FISH analyses performed to identify the appropriate "acrocentric" plant chromosome for *de novo* plant ACes formation.

15 **2. Bacteriophage lambda integrase-based site-specific recombination system**

An integral part of the platform technology includes a site-specific recombination system that allows the placement of selected gene targets or genomic fragments onto the platform chromosomes. Any such system
20 may be used. In particular, a method is provided for insertion of additional DNA fragments into the platform chromosome residing in the cell via sequence-specific recombination using the recombinase activity of the bacteriophage lambda integrase. The lambda integrase system is exemplary of the recombination systems contemplated for ACes. Any
25 known recombination system, including any described herein, particularly any that operates without the need for additional factors or that, by virtue of mutation, does not require additional factors, is contemplated.

-53-

As noted the lambda integrase system provided herein can be used with natural chromosomes and artificial chromosomes in addition to ACes. Single or a plurality of recombination sites, which may be the same or different, are introduced into artificial chromosomes to produce
5 artificial chromosome platforms.

3. Creation of bacteriophage lambda integrase site-specific recombination system

The lambda phage-encoded integrase (designated Int) is a prototypical member of the integrase family. Int effects integration and
10 excision of the phage in and out of the *E. coli* genome via recombination between pairs of attachment sites designated *attB/attP* and *attL/attR*. Each *att* site contains two inverted 9 base pair core Int binding sites and a 7 base pair overlap region that is identical in wild-type *att* sites. Each site, except for *attB* contains additional Int binding sites. In flanking
15 regions, there are recognition sequences for accessory DNA binding proteins, such as integration host factor (IHF), factor for inversion stimulation (FIS) and the phage encoded excision protein (XIS). Except for *attB*, Int is a heterobivalent DNA-binding protein and, with assistance from the accessory proteins and negative DNA supercoiling, binds
20 simultaneously to core and arm sites within the same *att* site.

Int, like Cre and FLP, executes an ordered sequential pair of strand exchanges during integrative and excisive recombination. The natural pairs of target sequences for Int, *attB* and *attP* or *attL* and *attR* are located on the same or different DNA molecules resulting in intra or
25 intermolecular recombination, respectively. For example, intramolecular recombination occurs between inversely oriented *attB* and *attP*, or between *attL* and *attR* sequences, respectively, leading to inversion of the intervening DNA segment.

-54-

Like the recombinase systems, such as Cre and FLP, Int directs site-specific recombination. Unlike the other systems, such as Cre and FLP, Int generally requires additional protein factors for integrative and excisive recombination and negative supercoiling for integrative recombination.

5 Hence, the Int system had not been used in eukaryotic targeting systems.

Mutant Int proteins, designated Int-h (E174K) and a derivative thereof Int-h/218(E174K/E218K) do not require accessory proteins to perform intramolecular integrative and excisive recombination in co-transfection assays in human cells (Lorbach *et al.* (2000) *J Mol. Biol.*
10 296:1175-1181); wild-type Int does not catalyze intramolecular recombination in human cells harboring target sites *attB* and *attP*. Hence it had been demonstrated that mutant Int can catalyze factor-independent recombination events in human cells.

There has been no demonstration by others that this system can be
15 used for engineering of eukaryotic genomes or chromosomes. Provided herein are chromosomes, including artificial chromosomes, such as but not limited to ACes that contain *att* sites (e.g., platform ACes), and the use of such chromosomes for targeted integration of heterologous DNA into such chromosomes in eukaryotic cells, including animal, such as
20 rodent and human, and plant cells. Mutant Int provided herein is shown to effect site-directed recombination between sites in artificial chromosomes and vectors containing cognate sites.

An additional component of the chromosome-based platform technology is the site-specific integration of target DNA sequences onto
25 the platform. For this the native bacteriophage lambda integrase has been modified to carry out this sequence specific DNA recombination event in eukaryotic cells. The bacteriophage lambda integrase and its cognate DNA substrate *att* is a member of the site-specific recombinase family that also includes the bacteriophage P1 Cre/lox system as well as

-55-

the *Saccharomyces cerevisiae* 2 micron based FLP/FRT system (see, e.g., Landy (1989) *Ann. Rev. Biochem* 58:913-949; Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398-3402; Broach *et al.* (1982) *Cell* 29:227-234).

5 By combining DNA endonuclease and DNA ligase activity these recombinases recognize and catalyze DNA exchanges between sequences flanking the recognition site. During the integration of lambda genome into the *E. coli* (lambda recombination) genome, the phage integrase (INT) in association with accessory proteins catalyzes the DNA exchange
10 between the attP site of the phage genome and the attB site of the bacterial genome resulting in the formation of attL and attR sites (Figure 6). The engineered bacteriophage lambda integrase has been produced herein to carry out an intermolecular DNA recombination event between
15 an incoming DNA molecule (primarily on a vector containing the bacterial attB site) and the chromosome-based platform carrying the lambda attP sequence independent of lambda bacteriophage or bacterial accessory proteins.

In contrast to the bi-directional Cre/lox and FLP/FRT system, the engineered lambda recombination system derived for chromosome-based
20 platform technology is advantageously unidirectional because accessory proteins, which are absent, are required for excision of integrated nucleic acid upon further exposure to the lambda Int recombinase.

4. Creation of platform chromosome containing single or multiple sequence-specific recombination sites

25 a. Multiple sites

For the creation of a platform chromosome containing multiple, sequence-specific recombination sites, artificial chromosomes are produced as depicted in Figure 5 and Example 3. As discussed above, artificial chromosomes can be produced using any suitable methodology,

-56-

including those described in U.S. Patent Nos. 5,288,625; 5,712,134; 5,891,691; 6,025,155. Briefly, to prepare artificial chromosomes containing multiple recombination (e.g., integration) sites, nucleic acid (either in the form a one or more plasmids, such as the plasmid pSV40193attPsensePUR set forth in Example 3) is targeted into an amplifiable region of a chromosome, such as the pericentric region of a chromosome. Among such regions are the rDNA gene loci in acrocentric mammalian chromosomes. Hence, targeting nucleic acid for integration into the rDNA region of mammalian acrocentric chromosomes can include the mouse rDNA fragments (for targeting into rodent cell lines) or large human rDNA regions on BAC/PAC vectors (or subclones thereof in standard vectors) for targeting into human acrocentric chromosomes, such as for human gene therapy applications. The targeting nucleic acid generally includes a detectable or selectable marker, such as antibiotic resistance, such as puromycin and hygromycin, a recombination site (such as attP, attB, attL, attR or the like), and/or human selectable markers as required for gene therapy applications. Cells are grown under conditions that result in amplification and ultimately production of ACes artificial chromosomes having multiple recombination (e.g., integration) sites therein. ACes having the desired size are selected for further engineering.

b. Creation of platform chromosome containing a single sequence-specific recombination site

In this method a mammalian platform artificial chromosome is generated containing a single sequence-specific recombination site. In the Example below, this approach is demonstrated using a puromycin resistance marker for selection and a mouse rDNA fragment for targeting into the rDNA locus on mouse acrocentric chromosomes. Other selection markers and targeting DNA sequences as desired and known to those of

-57-

skill in the art can be used. Additional resistance markers include genes conferring resistance to the antibiotics neomycin, blasticidin, hygromycin and zeocin. For applications, such as gene therapy in which potentially immunogenic responses are to be avoided, host, such as human, derived

5 selectable markers or markers detectable with monoclonal antibodies (MAb) followed by fluorescent activated cell sorting (FACS) can be used. Examples in this class include, but are not limited to: human nerve growth factor receptor (detection with MAb); truncated human growth factor receptor (detection with MAb); mutant human dihydrofolate reductase

10 (DHFR; detectable using a fluorescent methotrexate substrate); secreted alkaline phosphatase (SEAP; detectable with fluorescent substrate); thymidylate synthase (TS; confers resistance to fluorodeoxyuridine); human CAD gene (confers resistance to N-phosphonacetyl-L-aspartate (PALA)).

15 To construct a platform artificial chromosome with a single site, an ACes artificial chromosome (or other artificial chromosome of interest) can be produced containing a selectable marker. A single sequence specific recombination site is targeted onto ACes via homologous recombination. For this, DNA sequences containing the site-specific

20 recombination sequence are flanked with DNA sequences homologous to a selected sequence in the chromosome. For example, when using a chromosome containing rDNA or satellite DNA, such DNA can be used as homologous sequences to target the site-specific recombination sequence onto the chromosome. A vector is designed to have these homologous

25 sequences flanking the site-specific recombination site and, after the appropriate restriction enzyme digest to generate free ends of homology to the chromosome, the DNA is transfected into cells harboring the chromosome. After transfection and integration of the site-specific cassette, homologous recombination events onto the platform

-58-

chromosome are subcloned and identified, for example by screening single cell subclones via expression of resistance or a fluorescent marker and PCR analysis. In one embodiment, a platform artificial chromosome, such as a platform ACes, that contains a single copy of the recombination site is selected. Examples 2B and 2D exemplify the process, and Figure 3 provides a diagram depicting one method for the creation of a platform mammalian chromosome containing a single sequence-specific recombination site.

10 **5. Lambda integrase mediated recombination of target gene expression vector onto platform chromosome**

 The third component of the chromosome-based platform technology involves the use of target gene expression vectors carrying, for example, genes for gene therapy, genes for transgenic animal or plant production, and those required for cellular protein production of interest.

15 Using lambda integrase mediated site-specific recombination, or any other recombinase-mediated site-specific recombination, the target gene expression vectors are introduced onto the selected chromosome platform. The use of target gene expression vector permits use of the *de novo* generated chromosome-based platforms for a wide range of gene

20 targets. Furthermore, chromosome platforms containing multiple *attP* sites provides the opportunity to incorporate multiple gene targets onto a single platform, thereby providing for expression of multiple gene targets, including the expression of cellular and genetic regulatory genes and the expression of all or parts of metabolic pathways. In addition to

25 expressing small target genes, such as cDNA and hybrid cDNA/artificial intron constructs, the chromosome-based platform can be used for engineering and expressing large genomic fragments carrying target genes along with its endogenous genomic promoter sequences. This is of importance, for example, where the therapy requires precise cell specific

-59-

expression and in instances where expression is best achieved from genomic clones rather than cDNA clones. Figure 9 provides a diagram summarizing one embodiment of the chromosome-based technology.

A feature of the target gene expression vector that is of interest to
5 include is a promoterless marker gene, which as exemplified (see, Figure 9) contains an upstream *attB* site (marker 2 on Figure 9). The nucleic acid encoding the marker is not expressed unless it is placed downstream from a promoter sequence. Using the recombinase technology provided herein, such as the lambda integrase technology (λ INT_{E174R} on figure 8)
10 provided herein, site-specific recombination between the *attB* site on the vector and the promoter-*attP* site (in the "sense" orientation) on the chromosome-based platform results in the expression of marker 2 on the target gene expression vector, thereby providing a positive selection for the lambda INT mediated site-specific recombination event. Site-specific
15 recombination events on the chromosome-based platform versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR. Examples of suitable marker 2 genes, include, but are not limited to, genes that confer resistance to toxic compounds or antibiotics,
20 fluorescence activated cell sorting (FACS) sortable cell surface markers and various fluorescent markers. Examples of these genes include, but are not limited to, human L26a^R (human homolog of *Saccharomyces cerevisiae* CYH⁸ gene), neomycin, puromycin, blasticidin, CD24 (see, e.g., US Patents 5,804,177 and 6,074,836), truncated CD4, truncated low
25 affinity nerve growth factor receptor (LNGFR), truncated LDL receptor, truncated human growth hormone receptor, GFP, RFP, BFP.

The target gene expression vectors contain a gene (target gene) for expression from the chromosome platform. The target gene can be expressed using various constitutive or regulated promoter systems

-60-

across various mammalian species. For the expression of multiple target genes within the same target gene expression vector, the expression of the multiple targets can be coordinately regulated via viral-based or human internal ribosome entry site (IRES) elements (see, *e.g.*, Jackson *et al.* (1990) *Trends Biochem Sci.* 15: 477-83; Oumard *et al.* (2000) *Mol. Cell. Biol.* 20: 2755-2759). Furthermore, using IRES type elements linked to a downstream fluorescent marker, *e.g.*, green, red or blue fluorescent proteins (GFP, RFP, BFP) allows for the identification of high expressing clones from the integrated target gene expression vector.

- 10 In certain embodiments described herein, the promoterless marker can be transcriptionally downstream of the heterologous nucleic acid, wherein the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.
- 15 In addition, the selectable marker and the heterologous nucleic acid can be transcriptionally linked by the presence of a IRES between them. As set forth herein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.
- 20 Expression from the target gene expression vector integrated onto the chromosome-based platform can be further enhanced using genomic insulator/boundary elements. The incorporation of insulator sequences into the target gene expression vector helps define boundaries in chromatin structure and thus minimizes influence of chromatin position
- 25 effects/gene silencing on the expression of the target gene (Bell *et al.* (1999) *Current Opinion in Genetics and Development* 9:191-198; Emery *et al.* (2000) *Proc. Natl. Acad. Sci. U.S.A.* 97:9150-9155). Examples of insulator elements that can be included onto target gene expression vector in order to optimize expression include, but are not limited to:

-61-

- 1) chicken β -globin HS4 element (Prioleau *et al.* (1999) *EMBO J* 18: 4035-4048);
- 2) matrix attachment regions (MAR; see, *e.g.*, Ramakrishnan *et al.* (2000) *Mol Cell. Biol.* 20:868-877);
- 5 3) scaffold attachment regions (SAR; see, *e.g.*, Auten *et al.* (1999) *Human Gene Therapy* 10:1389-1399); and
- 4) universal chromatin opening elements (UCOE; WO/0005393 and WO/0224930)

The copy number of the target gene can be controlled by sequentially adding multiple target gene expression vectors containing the target gene onto multiple integration sites on the chromosome platform. Likewise, the copy number of the target gene can be controlled within an individual target gene expression vector by the addition of DNA sequences that promote gene amplification. For example, gene amplification can be induced utilizing the dihydrofolate reductase (DHFR) minigene with subsequent selection with methotrexate (see, *e.g.*, Schimke (1984) *Cell* 37:705-713) or amplification promoting sequences from the rDNA locus (see, *e.g.*, Wegner *et al.* (1989) *Nucl. Acids Res.* 17: 9909-9932).

20 **6. Platforms with other recombinase system sites**

A "double *lox*" targeting strategy mediated by Cre-recombinase (Bethke *et al.* (1997) *Nucl. Acids Res.* 25:2828-2834) can be used. This strategy employs a pair of heterospecific *lox* sites—*loxA* and *loxB*, which differ by one nucleotide in the 8 bp spacer region. Both sites are engineered into the artificial chromosome and also onto the targeting DNA vector. This allows for a direct site-specific insertion of a commercially relevant gene or genes by a Cre-catalyzed double crossover event. In essence a platform ACes is engineered with a hygromycin-resistance gene flanked by the double *lox* sites generating *lox-ACes*, which is maintained

-62-

in the thymidine kinase deficient cell, LMtk(-). The gene of interest, for example, for testing purposes, the green fluorescence protein gene, GFP and a HSV thymidine kinase gene (tk) marker, are engineered between the appropriate *lox* sites of the targeting vector. The vector DNA is

5 cotransfected with plasmid pBS185 (Life Technologies) encoding the *Cre* recombinase gene into mammalian cells maintaining the dual-*lox* artificial chromosome. Transient expression of the *Cre* recombinase catalyzes the site-specific insertion of the gene and the tk-gene onto the artificial chromosome. The transfected cells are grown in HAT medium that

10 selects for only those cells that have integrated and expressed the thymidine kinase gene. The HAT^R colonies are screened by PCR analyses to identify artificial chromosomes with the desired insertion.

To generate the *lox-ACes*, Lambda-Hyg^R-*lox* DNA is transfected into the LMtk(-) cell line harboring the precursor *ACes*. Hygromycin-

15 resistant colonies are analyzed by FISH and Southern blotting for the presence of a single copy insert on the *ACes*.

To demonstrate the gene replacement technology, cell lines containing candidate *lox-ACes* are cotransfected with pTK-GFP-*lox* and pBS185 (encoding the *Cre* recombinase gene) DNA. After transfection,

20 transient expression of plasmid pBS185 will provide sufficient burst of *Cre* recombinase activity to catalyze DNA recombination at the *lox* sites. Thus, a double crossover event between the *ACes* target and the exogenous targeting plasmid carrying the *loxA* and *loxB* permits the simple replacement of the hygromycin-resistance gene on the *lox-ACes*

25 for the tk-GFP cassette from the targeting plasmid, with no integration of vector DNA. Transfected cells are grown in HAT-media to select for tk-expression. Correct targeting will result in the generation of HAT^R, hygromycin sensitive, and green fluorescent cells. The desired integration event is verified by Southern and PCR analyses. Specific PCR primer sets

-63-

are used to amplify DNA sequences flanking the individual *loxA* and *loxB* sites on the *lox-ACes* before and after homologous recombination.

D. Exemplary applications of the Platform *ACes*

Platform *ACes* are applicable and tractable for different/optimized
5 cell lines. Those that include a fluorescent marker, for example, can be purified and isolated using fluorescent activated cell sorting (FACS), and subsequently delivered to a target cell. Those with selectable markers provide for efficient selection and provide a growth advantage. Platform *ACes* allow multiple payload delivery of donor target vectors via a
10 positive-selection site-specific, recombination system, and they allow for the inclusion of additional genetic factors that improve protein production and protein quality.

The construction and use of the platform *ACes* as provided for each application may be similarly applied to other applications. Particular
15 descriptions are for exemplification.

1. Cellular Protein Production Platform *ACes* (CPP *ACes*)

As described herein, *ACes* can be produced from acrocentric chromosomes in rodent (mouse, hamster) cell lines via megareplicator induced amplification of heterochromatin/rDNA sequences. Such *ACes*
20 are ideal for cellular protein production as well as other applications described herein and known to those of skill in the art. *ACes* platforms that contain a plurality of recombination sites are particularly suitable for engineering as cellular protein production systems.

In one embodiment, CPP *ACes* involve a two-component system:
25 the platform chromosome containing multiple engineering sites and the donor target vector containing a platform-specific recombination site with designed expression cassettes (see Figure 9).

The platform *ACes* can be produced from any artificial chromosome, particularly the amplification-based artificial chromosomes.

-64-

For exemplification, they are produced from rodent artificial chromosomes produced from acrocentric chromosomes using the technology of U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183, in which nucleic acid is targeted to the
5 pericentric heterochromatic, and, particularly into rDNA to initiate the replication event(s). The ACes can be produced directly in the chosen cellular protein production cell lines, such as, but not limited to, CHO cells, hybridomas, plant cells, plant tissues, plant protoplasts, stem cells and plant calli.

10 a. **Platform Construction**

In the exemplary embodiment, the initial *de novo* platform construction requires co-transfecting with excess targeting DNA, such as, rDNA or lambda DNA without an *attP* region, and an engineered selectable marker. The engineered selectable marker should contain
15 promoter, generally a constitutive promoter, such as human, viral, i.e., adenovirus or SV40 promoter, including the human ferritin heavy chain promoter (SEQ ID NO:128), SV40 and EF1 α promoters, to control expression of a marker gene that provides a selective growth advantage to the cell. An example of such a marker gene is the *E. coli hisD* gene
20 (encoding histidinol dehydrogenase) which is homologous and analogous to the *S. typhimurium hisD* a dominant marker selection system for mammalian cells previously described (see, Hartman *et al.* (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85:8047-8051). Since histidine is an essential amino acid in mammals and a nutritional requirement in cell culture, the *E.*
25 *coli hisD* gene can be used to select for histidine prototrophy in defined media. Furthermore more stringent selection can be placed on the cells by including histinol in the medium. Histidinol is itself permeable and toxic to cells. The *hisD* provides a means of detoxification.

-65-

Placed between the promoter and the marker gene is the bacteriophage lambda *attP* site to use the bacteriophage lambda integrase dependent site-specific recombination system (described herein). The insertion of an *attP* site downstream of a promoter element provide
5 forward selection of site-specific recombination events onto the platform *ACes*.

b. Donor Target Vector Construction

A second component of the CPP platform *ACes* system involves the construction of donor target vectors containing a gene product(s) of
10 interest for the CPP platform *ACes*. Individual donor target vectors can be designed for each gene product to be expressed thus enabling maximum usage of a *de novo* constructed platform *ACes*, so that one or a few CPP platform *ACes* will be required for many gene targets.

A key feature of the donor vector target is the *promoterless* marker
15 gene containing an upstream *attB* site (marker 2 on figure 9). Normally the marker would not be expressed unless it is placed downstream of a promoter sequence. As discussed above, using the lambda integrase technology (λ INT_{E174R} on Figure 8 and Figure 9), site-specific recombination between the *attB* site on the vector and the promoter-*attP*
20 site on the CPP platform *ACes* result in the expression of the donor target vector marker providing positive selection for the site-specific event. Site-specific recombination events on the CPP *ACes* versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR.
25 In addition, since the lambda integrase reaction is unidirectional, i.e. excision reaction is not possible, a number of unique targets can be loaded onto the CPP platform *ACes* limited only by the number of markers available.

-66-

Additional features of the donor target vector include gene target expression cassettes flanked by either chromatin insulator regions, matrix attachment regions (MAR) or scaffold attachment regions (SAR). The use of these regions will provide a more "open" chromatin environment for gene expression and help alleviate silencing. An example of such a cassette for expressing a monoclonal antibody is described. For this purpose, a strong constitutive promoter, e.g. chicken β -actin or RNA Poll, is used to drive the expression of the heavy and light chain open reading frames. The heavy and light chain sequences flank a nonattenuated human IRES (IRES_H; from the 5'UTR of NRF1 gene; see Oumard et al., 2000, *Mol. and Cell Biol.*, 20(8):2755-2759) element thereby coordinating transcription of both heavy and light chain sequence. Distal to the light chain open reading frame resides an additional viral encoded IRES (IRES_V modified ECMV internal ribosomal entry site (IRES)) element attenuating the expression of the fluorescent marker gene hrGFP from *Renilla* (Stratagene). By linking the hrGFP with an attenuated IRES, the heavy and light chains along with the hrGFP are monocistronic. Thus, the identification of hrGFP fluorescing cells will provide a means to detect protein producing cells. In addition, high producing cell lines can be identified and isolated by FACS thereby decreasing the time frame in finding high expressers. Functional monoclonal antibody will be confirmed by ELISA.

c. Additional components in cellular protein production platform ACes (CPP ACes)

In addition to the aforementioned CPP ACes system, other genetic factors can be included to enhance the yield and quality of the expressed protein. Again to provide maximum flexibility, these additional factors can be inserted onto the CPP platform ACes by λ INTE174R dependent site-specific recombination. Other factors that could be used with a CPP

-67-

Platform ACes include for example, adenovirus E1a transactivation system which upregulates both cellular and viral promoters (see, e.g., Svensson and Akusjarvi (1984) EMBO 3:789-794; and US patents 5,866,359; 4,775,630 and 4,920,211).

5 d. **Targets for CHO-ACes engineering to enhance cell growth, such as CHO cell growth and protein production/ quality**

 If adding these additional factors onto the CPP ACes is not prudent or desired, the host cell, CHO cells, can be engineered to express these
10 factors (see, below, targets for CHO-ACes engineering to enhance CHO cell growth and protein production/quality). Additional factors to consider including are addition of insulin or IGF-1 to sustain viability; human sialyltransferases or related factors to produce more human-like glycoproteins; expression of factors to decrease ammonium accumulation
15 during cell growth; expression of factors to inhibit apoptosis; expression of factors to improve protein secretion and protein folding; and expression of factors to permit serum-free transfection and selection.

 1) **Addition of insulin or IGF-1 to sustain viability**

20 Stimulatory factors and/or their receptors are expressed to set up an autocrine loop, to improve cell growth, such as CHO cell growth. Two exemplary candidates are insulin and IGF-1 (see, Biotechnol Prog 2000 Sep;16(5):693-7). Insulin is the most commonly used growth factor for sustaining cell growth and viability in serum-free Chinese hamster ovary
25 (CHO) cell cultures. Insulin and IGF-1 analog (LongR(3) serve as growth and viability factors for CHO cells.

 CHO cells were modified to produce higher levels of essential nutrients and factors. A serum-free (SF) medium for dihydrofolate reductase-deficient Chinese hamster ovary cells (DG44 cells) was
30 prepared. Chinese hamster ovary cells (DG44 cells), which are normally

-68-

maintained in 10% serum medium, were gradually weaned to 0.5% serum medium to increase the probability of successful growth in SF medium (see, Kim *et al.* (199) *In Vitro Cell Dev Biol Anim* 35(4):178-82). A SF medium (SF-DG44) was formulated by supplementing the basal
5 medium with these components; basal medium was prepared by supplementing Dulbecco's modified Eagle's medium and Ham's nutrient mixture F12 with hypoxanthine (10 mg/l) and thymidine (10 mg/l). Development of a SF medium for DG44 cells was facilitated using a Plackett-Burman design technique and weaning of cells.

10

2) Human sialyltransferases or related factors to produce more human-like glycoproteins

CHO cells have been modified by increasing their ability to process
15 protein via addition of complex carbohydrates. This has been achieved by overexpression of relevant processing enzymes, or in some cases, reducing expression of relevant enzymes (see, Bragonzi *et al.* (2000) *Biochim Biophys Acta* 1474(3):273-282; see, also Weikert *et al.* (1999) *Nature biotech.* 17:1116-11121; Ferrari J *et al.* (1998) *Biotechnol Bioeng*
20 60(5):589-95). A CHO cell line expressing alpha2,6-sialyltransferase was developed for the production of human-like sialylated recombinant glycoproteins. The sialylation defect of CHO cells can be corrected by transfecting the alpha2,6-sialyltransferase (alpha2,6-ST) cDNA into the cells. Glycoproteins produced by such CHO cells display alpha2,6-and
25 alpha2,3-linked terminal sialic acid residues, similar to human glycoproteins.

As another example for improving the production of human-like sialylated recombinant glycoproteins, a CHO cell line has been developed that constitutively expresses sialidase antisense RNA (see, Ferrari J *et al.*
30 (1998) *Biotechnol Bioeng* 60(5):589-95). Several antisense expression

-69-

vectors were prepared using different regions of the sialidase gene. Co-transfection of the antisense constructs with a vector conferring puromycin resistance gave rise to over 40 puromycin resistant clones that were screened for sialidase activity. A 5' 474 bp coding segment of the
5 sialidase cDNA, in the inverted orientation in an SV 40-based expression vector, gave maximal reduction of the sialidase activity to about 40% wild-type values.

Oligosaccharide biosynthesis pathways in mammalian cells have been engineered for generation of recombinant glycoproteins (see, *e.g.*,
10 Sburlati (1998) *Biotechnol Prog* 14(2):189-92), which describes a Chinese hamster ovary (CHO) cell line capable of producing bisected oligosaccharides on glycoproteins. This cell line was created by overexpression of a recombinant N-acetylglucosaminyltransferase III (GnT-III) (see, also, Prati *et al.* (1998) *Biotechnol Bioeng* 59(4):445-50, which
15 describes antisense strategies for glycosylation engineering of CHO cells).

3) Expression of factors to decrease ammonium accumulation during cell growth

Excess ammonium, which is a by-product of CHO cell metabolism
20 can have detrimental effects on cell growth and protein quality (see, Yang *et al.* (2000) *Biotechnol Bioeng* 68(4):370-80). To solve this problem ammonium levels were modified by overexpressing carbamoyl phosphate synthetase I and ornithine transcarbamoylase or glutamine synthetase in CHO cells. Such modification resulted in reduced ammonium levels
25 observed and an increase in the growth rate (see Kim *et al.* (2000) *J Biotechnol* 81(2-3):129-40; and Enosawa *et al.* (1997) *Cell Transplant* 6(5):537-40).

4) Expression of factors to improve protein secretion and protein folding

-70-

Overexpression of relevant enzymes can be engineered into the ACes to improve protein secretion and folding.

5) Expression of factors to permit serum-free transfection and selection

5 It is advantageous to have the ability to convert CHO cells in suspension growing in serum free medium to adherence with out having to resort to serum addition. Laminin or fibronectin addition is sufficient to make cells adherent (see, *e.g.*, Zaworski *et al.* (1993) *Biotechniques* 15(5):863-6) so that expressing either of these genes in CHO cells under
10 an inducible promoter should allow for reversible shift to adherence without requiring serum addition.

2. Platform ACes and Gene Therapy

The platform ACes provided herein are contemplated for use in mammalian gene therapy, particularly human gene therapy. Human ACes
15 can be derived from human acrocentric chromosomes from human host cells, in which the amplified sequences are heterochromatic and/or human rDNA. Different platform ACes applicable for different tissue cell types are provided. The ACes for gene therapy can contain a single copy of a therapeutic gene inserted into a defined location on platform ACes.
20 Therapeutic genes include genomic clones, cDNA, hybrid genes and other combinations of sequences. Preferred selectable markers are those from the mammalian host, such as human derived factors so that they are non-immunogenic, non-toxic and allow for efficient selection, such as by FACS and/or drug resistance.

25 Platform ACes, useful for gene therapy and other applications, as noted herein, can be generated by megareplicator dependent amplification, such as by the methods in U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183. In one embodiment, human ACes are produced using

-71-

human rDNA constructs that target rDNA arrays on human acrocentric chromosomes and induce the megareplicator in human cells, particularly in primary cell lines (with sufficient number of doublings to form the ACes) or stem cells (such as hematopoietic stem cells, mesenchymal stem cells, adult stem cells or embryonic stem cells) to avoid the introduction of potentially harmful rearranged DNA sequences present in many transformed cell lines. Megareplicator induced ACes formation can result in multiple copies of targeting DNA/selectable markers in each amplification block on both chromosomal arms of the platform ACes.

10 In view of the considerations regarding immunogenicity and toxicity, the production of human platform ACes for gene therapy applications employs a two component system analogous to the platform ACes designed for cellular protein production (CPP platform ACes). The system includes a platform chromosome of entirely human DNA origin

15 containing multiple engineering sites and a gene target vector carrying the therapeutic gene of interest.

a. Platform Construction

The initial *de novo* construction of the platform chromosome employs the co-transfection of excess targeting DNA and a selectable

20 marker. In one embodiment, the DNA is targeted to the rDNA arrays on the human acrocentric chromosomes (chromosomes 13, 14, 15, 21 and 22). For example, two large human rDNA containing PAC clones 18714 and 18720 and the human PAC clone 558F8 are used for targeting (Genome Research (ML) now Incyte, BACPAC Resources, 747 52nd

25 Street, Oakland CA). The mouse rDNA clone pFK161 (SEQ ID NO: 118), which was used to make the human SATAC from the 94-3 hamster/human hybrid cell line (see, *e.g.*, published International PCT application No. WO 97/40183 and Csonka, *et al*, *Journal of Cell Science*

-72-

113:3207-32161 and Example 1 for a description of pFK161) can also be used.

For animal applications, selectable markers should be non-immunogenic in the animal, such as a human, and include, but are not limited to: human nerve growth factor receptor (detected with a MAb, such as described in US patent 6,365,373); truncated human growth factor receptor (detected with MAb), mutant human dihydrofolate reductase (DHFR; fluorescent MTX substrate available); secreted alkaline phosphatase (SEAP; fluorescent substrate available); human thymidylate synthase (TS; confers resistance to anti-cancer agent fluorodeoxyuridine); human glutathione S-transferase alpha (GSTA1; conjugates glutathione to the stem cell selective alkylator busulfan; chemoprotective selectable marker in CD34+ cells); CD24 cell surface antigen in hematopoietic stem cells; human CAD gene to confer resistance to N-phosphonacetyl-L-aspartate (PALA); human multi-drug resistance-1 (MDR-1; P-glycoprotein surface protein selectable by increased drug resistance or enriched by FACS); human CD25 (IL-2 α ; detectable by Mab-FITC); Methylguanine-DNA methyltransferase (MGMT; selectable by carmustine); and Cytidine deaminase (CD; selectable by Ara-C).

Since megareplicator induced amplification generates multiple copies of the selectable marker, a second consideration for the selection of the human marker is the resulting dose of the expressed marker after ACes formation. High level of expression of certain markers may be detrimental to the cell and/or result in autoimmunity. One method to decrease the dose of the marker protein is by shortening its half-life, such as via the fusion of the well-conserved human ubiquitin tag (a 76 amino acid sequence) thus leading to increased turnover of the selectable marker. This has been used successfully for a number of reporter

-73-

systems including DHFR (see, *e.g.*, Stack *et al.* (2000) *Nature Biotechnology* 18:1298-1302 and references cited therein).

Using the ubiquitin tagged protein, a human selectable marker system analogous to the CPP ACes described herein is constructed.

5 Briefly, a tagged selectable marker, such as for example one of those described herein, is cloned downstream of an *attP* site and expressed from a human promoter. Exemplary promoters contemplated for use herein include, but are not limited to, the human ferritin heavy chain promoter (SEQ ID NO:128); RNA Poll; EF1 α ; TR; glyceraldehyde-3-
10 phosphate dehydrogenase core promoter (GAP); a GAP core promoter including a proximal insulin inducible element the intervening GAP sequence; phosphofructokinase promoter; and phosphoglycerate kinase promoter. Also contemplated herein is an aldolase A promoter H1 & H2 (representing closely spaced transcriptional start sites) along with the
15 proximal H enhancer. There are 4 promoters (*e.g.*, transcriptional start sites) for this gene, each having different regulatory and tissue activity. The H (most proximal 2) promoters are ubiquitously expressed off the H enhancer. This resulting marker can then be co-transfected along with excess human rDNA targeting sequence into the host cells. An important
20 criteria for the selection of the

recipient cells is sufficient number of cell doublings for the formation and detection of ACes. Accordingly, the co-transfections should be attempted in human primary cells that can be cultured for long periods of
25 time, such as for example, stem cells (*e.g.*, hematopoietic, mesenchymal, adult or embryonic stem cells), or the like. Additional cell types, include, but are not limited to: single gene transfected cells exhibiting increased life-span; over-expressing c-myc cells, *e.g.* MSU1.1 (Morgan *et al.*, 1991, *Exp. Cell Res.*, Nov;197(1):125-136); over-expressing telomerase lines,

-74-

such as TERT cells; SV40 large T-antigen transfected lines; tumor cell lines, such as HT1080; and hybrid human cell lines, such as the 94-3 hamster/human hybrid cell line.

b. Gene Target Vector

5 The second component of the GT platform *ACes* (GT *ACes*) system involves the use of engineered target vectors carrying the therapeutic gene of interest. These are introduced onto the GT platform *ACes* via site-specific recombination. As with the CPP *ACes*, the use of engineered target vectors maximizes the use of the *de novo* generated GT platform
10 *ACes* for most gene targets. Furthermore, using lambda integrase technology, GT platform *ACes* containing multiple *attP* sites permits the opportunity to incorporate multiple therapeutic targets onto a single platform. This could be of value in cases where a defined therapy requires multiple gene targets, a single therapeutic target requires an
15 additional gene regulatory factor or a GT *ACes* requires a "kill" switch.

Similar to the CPP *ACes*, a feature of the gene target vector is the *promoterless* marker gene containing an upstream *attB* site (marker 2 on Figure 9). Normally, the marker (in this case, a cell surface antigen that can be sorted by FACS would be ideal) would not be expressed unless it
20 is placed downstream of a promoter sequence. Using the lambda integrase technology (λ INT_{E174R} on figure 9), site-specific recombination between the *attB* site on the vector and the promoter- *attP* site on the GT platform *ACes* results in the expression of marker#2 on the gene target vector, i.e. positive selection for the site-specific event. Site-specific
25 recombination events on the GT *ACes* versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR.

For expression of the therapeutic gene, human specific promoters, such as a ferritin heavy chain promoter (SEQ ID NO:128); EF1 α or RNA

-75-

Poll, are used. These promoters are for high level expression of a cDNA encoded therapeutic protein. In addition to expressing cDNA (or even hybrid cDNA/artificial intron constructs), the GT platform *ACes* are used for engineering and expressing large genomic fragments carrying
5 therapeutic genes of interest expressed from native promoter sequences. This is of importance in situations where the therapy requires precise cell specific expression or in instances where expression is best achieved from genomic clones versus cDNA.

10 **3. Selectable markers for use, for example, in Gene Therapy (GT)**

The following are selectable markers that can be incorporated into human *ACes* and used for selection.

15 **Dual Resistance to 4-Hydroperoxycyclophosphamide and Methotrexate by Retroviral Transfer of the Human Aldehyde Dehydrogenase Class 1 Gene and a Mutated Dihydrofolate Reductase Gene**

The genetic transfer of drug resistance to hematopoietic cells is one approach to overcoming myelosuppression caused by high-dose chemotherapy. Because cyclophosphamide (CTX) and methotrexate
20 (MTX) are commonly used non-cross-resistant drugs, generation of dual drug resistance in hematopoietic cells that allows dose intensification may increase anti-tumor effects and circumvent the emergence of drug-resistant tumors, a retroviral vector containing a human cytosolic ALDH-1-encoding DNA clone and a human doubly mutated DHFR-encoding
25 clone (Phe22/Ser31; termed F/S in the description of constructs) to generate increased resistance to CTX and MTX were constructed (Takebe *et al.* (2001) *Mol Ther* 3(1):88-96). This construct may be useful for protecting patients from high-dose CTX- and MTX-induced myelosuppression. *ACes* can be similarly constructed.

-76-

5 **Multiple mechanisms of N-phosphonacetyl-L-aspartate resistance in human cell lines: carbamyl-P synthetase/aspartate transcarbamylase/dihydro-orotase gene amplification is frequent only when chromosome 2 is rearranged**

 Rodent cells resistant to N-phosphonacetyl-L-aspartate (PALA) invariably contain amplified carbamyl-P synthetase/aspartate transcarbamylase/dihydro-orotase (CAD) genes, usually in widely spaced tandem arrays present as extensions of the same chromosome arm that
10 carries a single copy of CAD in normal cells (Smith *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:1816-21). In contrast, amplification of CAD is very infrequent in several human tumor cell lines. Cell lines with minimal chromosomal rearrangement and with unrearranged copies of chromosome 2 rarely develop intrachromosomal amplifications of CAD.
15 These cells frequently become resistant to PALA through a mechanism that increases the aspartate transcarbamylase activity with no increase in CAD copy number, or they obtain one extra copy of CAD by forming an isochromosome 2p or by retaining an extra copy of chromosome 2. In cells with multiple chromosomal aberrations and rearranged copies of
20 chromosome 2, amplification of CAD as tandem arrays from rearranged chromosomes is the most frequent mechanism of PALA resistance. All of these different mechanisms of PALA resistance are blocked in normal human fibroblasts. Thus, ACes with multiple copies of the CAD gene would provide PALA resistance.

25 **Retroviral coexpression of thymidylate synthase and dihydrofolate reductase confers fluoropyrimidine and antifolate resistance**

 Retroviral gene transfer of dominant selectable markers into hematopoietic cells can be used to select genetically modified cells in vivo
30 or to attenuate the toxic effects of chemotherapeutic agents. Fantz *et al.* ((1998) *Biochem Biophys Res Comm* 243(1):6-12) have shown that

-77-

retroviral gene transfer of thymidylate synthase (TS) confers resistance to TS directed anticancer agents and that co-expression of TS and dihydrofolate reductase (DHFR) confers resistance to TS and DHFR cytotoxic agents. Retroviral vectors encoding *Escherichia coli* TS, human TS, and the Tyr-to-His at residue 33 variant of human TS (Y33HhTS) were constructed and fibroblasts transfected with these vectors conferred comparable resistance to the TS-directed agent fluorodeoxyuridine (FdUrd, approximately 4-fold). Retroviral vectors that encode dual expression of Y33HhTS and the human L22Y DHFR (L22YhDHFR) variants conferred resistance to FdUrd (3- to 5-fold) and trimetrexate (30- to 140-fold). A L22YhDHFR-Y33HhTS chimeric retroviral vector was also constructed and transduced cells were resistant to FdUrd (3-fold), AG337 (3-fold), trimetrexate (100-fold) and methotrexate (5-fold). These results show that recombinant retroviruses can be used to transfer the cDNA that encodes TS and DHFR and dual expression in transduced cells is sufficiently high to confer resistance to TS and DHFR directed anticancer agents. ACes can be similarly constructed.

Human CD34+ cells do not express glutathione S-transferases alpha

The expression of glutathione S-transferases alpha (GST alpha) in human hematopoietic CD34+ cells and bone marrow was studied using RT-PCR and immunoblotting (Czerwinski M, Kiem *et al.* (1997) *Gene Ther* 4(3):268-70). The GSTA1 protein conjugates glutathione to the stem cell selective alkylator busulfan. This reaction is the major pathway of elimination of the compound from the human body. Human hematopoietic CD34+ cells and bone marrow do not express GSTA1 message, which was present at a high level in liver, an organ relatively resistant to busulfan toxicity in comparison to bone marrow. Similarly, baboon CD34+ cells and dog bone marrow do not express GSTA1. Thus, human

-78-

GSTA1 is a chemoprotective selectable marker in human stem cell gene therapy and could be employed in ACes construction.

Selection of retrovirally transduced hematopoietic cells using CD24 as a marker of gene transfer

5 Pawliuk *et al.* ((1994) *Blood* 84(9):2868-2877) have investigated the use of a cell surface antigen as a dominant selectable marker to facilitate the detection and selection of retrovirally infected target cells. The small coding region of the human cell surface antigen CD24 (approximately 240 bp) was introduced into a myeloproliferative sarcoma virus (MPSV)-based retroviral vector, which was then used to infect day 4
10 5-fluorouracil (5-FU)-treated murine bone marrow cells. Within 48 hours of termination of the infection procedure CD24-expressing cells were selected by fluorescent-activated cell sorting (FACS) with an antibody directed against the CD24 antigen. Functional analysis of these cells
15 showed that they included not only *in vitro* clonogenic progenitors and day 12 colony-forming unit-spleen but also cells capable of competitive long-term hematopoietic repopulation. Double-antibody labeling studies performed on recipients of retrovirally transduced marrow cells showed that some granulocytes, macrophages, erythrocytes, and, to a lesser
20 extent, B and T lymphocytes still expressed the transduced CD24 gene at high levels 4 months later. No gross abnormalities in hematopoiesis were detected in mice repopulated with CD24-expressing cells. These results show that the use of the CD24 cell surface antigen as a retrovirally encoded marker permits rapid, efficient, and nontoxic selection *in vitro* of
25 infected primary cells, facilitates tracking and phenotyping of their progeny, and provides a tool to identify elements that regulate the expression of transduced genes in the most primitive hematopoietic cells. ACes could be similarly constructed.

-79-

DeltahGHR, a biosafe cell surface-labeling molecule for analysis and selection of genetically transduced human cells

A selectable marker for retroviral transduction and selection of human and murine cells is known (see, Garcia-Ortiz *et al.* (2000) *Hum Gene Ther* 11(2):333-46). The molecule expressed on the cell surface of the transduced population is a truncated version of human growth hormone receptor (deltahGHR), capable of ligand (hGH) binding, but devoid of the domains involved in signal triggering. The engineered molecule is stably expressed in the target cells as an inert protein unable to trigger proliferation or to rescue the cells from apoptosis after ligand binding. This new marker, has a wide application spectrum, since hGHR in the human adult is highly expressed only in liver cells, and lower levels have been reported in certain lymphocyte cell populations. The deltahGHR label has high biosafety potential, as it belongs to a well-characterized hormonal system that is nonessential in adults, and there is extensive clinical experience with hGH administration in humans. The differential binding properties of several monoclonal antibodies (MAbs) are used in a cell rescue method in which the antibody used to select deltahGHR-transduced cells is eluted by competition with hGH or, alternatively biotinylated hGH is used to capture tagged cells. In the latter system, the final purified population is recovered free of attached antibodies in hGH (a substance approved for human use)-containing medium. Such a system could be used to identify ACes containing cells.

4. Transgenic models for evaluation of genes and discovery of new traits in plants

Of interest is the use of plants and plant cells containing artificial chromosomes for the evaluation of new genetic combinations and discovery of new traits. Artificial chromosomes, by virtue of the fact that they can contain significant amounts of DNA can also therefore encode

-80-

numerous genes and accordingly a multiplicity of traits. It is contemplated here that artificial chromosomes, when formed from one plant species, can be evaluated in a second plant species. The resultant phenotypic changes observed, for example, can indicate the nature of the genes contained within the DNA contained within the artificial chromosome, and hence permit the identification of novel genetic activities. Artificial chromosomes containing euchromatic DNA or partially containing euchromatic DNA can serve as a valuable source of new traits when transferred to an alien plant cell environment. For example, it is contemplated that artificial chromosomes derived from dicot plant species can be introduced into monocot plant species by transferring a dicot artificial chromosome. The dicot artificial chromosome possessing a region of euchromatic DNA containing expressed genes.

The artificial chromosomes can be designed to allow the artificial chromosome to recombine with the naturally occurring plant DNA in such a fashion that a large region of naturally occurring plant DNA becomes incorporated into the artificial chromosome. This allows the artificial chromosome to contain new genetic activities and hence carry novel traits. For example, an artificial chromosome can be introduced into a wild relative of a crop plant under conditions whereby a portion of the DNA present in the chromosomes of the wild relative is transferred to the artificial chromosome. After isolation of the artificial chromosome, this naturally occurring region of DNA from the wild relative, now located on the artificial chromosome can be introduced into the domesticated crop species and the genes encoded within the transferred DNA expressed and evaluated for utility. New traits and gene systems can be discovered in this fashion. The artificial chromosome can be modified to contain sequences that promote homologous recombination within plant cells, or

-81-

be modified to contain a genetic system that functions as a site-specific recombination system.

Artificial chromosomes modified to recombine with plant DNA offer many advantages for the discovery and evaluation of traits in different plant species. When the artificial chromosome containing DNA from one plant species is introduced into a new plant species, new traits and genes can be introduced. This use of an artificial chromosome allows for the ability to overcome the sexual barrier that prevents transfer of genes from one plant species to another species. Using artificial chromosomes in this fashion allows for many potentially valuable traits to be identified including traits that are typically found in wild species. Other valuable applications for artificial chromosomes include the ability to transfer large regions of DNA from one plant species to another, such as DNA encoding potentially valuable traits such as altered oil, carbohydrate or protein composition, multiple genes encoding enzymes capable of producing valuable plant secondary metabolites, genetic systems encoding valuable agronomic traits such as disease and insect resistance, genes encoding functions that allow association with soil bacterium such as growth promoting bacteria or nitrogen fixing bacteria, or genes encoding traits that confer freezing, drought or other stress tolerances. In this fashion, artificial chromosomes can be used to discover regions of plant DNA that encode valuable traits.

The artificial chromosome can also be designed to allow the transfer and subsequent incorporation of these valuable traits now located on the artificial chromosome into the natural chromosomes of a plant species. In this fashion the artificial chromosomes can be used to transfer large regions of DNA encoding traits normally found in one plant species into another plant species. In this fashion, it is possible to derive a plant cell that no longer needs to carry an artificial chromosome to

-82-

posses the novel trait. Thus, the artificial chromosome would serve as the transfer mechanism to permit the formation of plants with greater degree of genetic diversity.

The design of an artificial chromosome to accomplish the afore-
5 mentioned purposes can include within the artificial chromosome the presence of specific DNA sequences capable of acting as sites for homologous recombination to take place. For example, the DNA sequence of *Arabidopsis* is now known. To construct an artificial chromosome capable of recombining with a specific region of *Arabidopsis*
10 DNA, a sequence of *Arabidopsis* DNA, normally located near a chromosomal location encoding genes of potential interest can be introduced into an artificial chromosome by methods provided herein. It may be desirable to include a second region of DNA within the artificial chromosome that provides a second flanking sequence to the region
15 encoding genes of potential interest, to promote a double recombination event which would ensure transfer of the entire chromosomal region, encoding genes of potential interest, to the artificial chromosome. The modified artificial chromosome, containing the DNA sequences capable of homologous recombination region, can then be introduced into
20 *Arabidopsis* cells and the homologous recombination event selected.

It is convenient to include a marker gene to allow for the selection of a homologous recombination event. The marker gene is preferably inactive unless activated by an appropriate homologous recombination event. For example, US 5,272,071, describes a method where an
25 inactive plant gene is activated by a recombination event such that desired homologous recombination events can be easily scored. Similarly, US 5,501,967 describes a method for the selection of homologous recombination events by activation of a silent selection gene first introduced into the plant DNA, the gene being activated by an appropriate

-83-

homologous recombination event. Both of these methods can be applied to enable a selective process to be included to select for recombination between an artificial chromosome and a plant chromosome. Once the homologous recombination event is detected, the artificial chromosome,
5 once selected, is isolated and introduced into a recipient cell, for example, tobacco, corn, wheat or rice, and the expression of the newly introduced DNA sequences evaluated.

Phenotypic changes in the recipient plant cells containing the artificial chromosome, or in regenerated plants containing the artificial
10 chromosome, allows for the evaluation of the nature of the traits encoded by the *Arabidopsis* DNA, under conditions naturally found in plant cells, including the naturally occurring arrangement of DNA sequences responsible for the developmental control of the traits in the normal chromosomal environment.

15 Traits such as durable fungal or bacterial disease resistance, new oil and carbohydrate compositions, valuable secondary metabolites such as phytosterols, flavonoids, efficient nitrogen fixation or mineral utilization, resistance to extremes of drought, heat or cold are all found within different populations of plant species and are often governed by
20 multiple genes. The use of single gene transformation technologies does not permit the evaluation of the multiplicity of genes controlling many valuable traits. Thus, incorporation of these genes into artificial chromosomes allows the rapid evaluation of the utility of these genetic combinations in heterologous plant species.

25 The large scale order and structure of the artificial chromosome provides a number of unique advantages in screening for new utilities or novel phenotypes within heterologous plant species. The size of new DNA that can be carried by an artificial chromosome can be millions of base pairs of DNA, representing potentially numerous genes that may

-84-

have novel utility in a heterologous plant cell. The artificial chromosome is a "natural" environment for gene expression, the problems of variable gene expression and silencing seen for genes transferred by random insertion into a genome should not be observed. Similarly, there is no
5 need to engineer the genes for expression, and the genes inserted would not need to be recombinant genes. Thus, one expects the expression from the transferred genes to be temporal and spatial, as observed in the species from where the genes were initially isolated. A valuable feature for these utilities is the ability to isolate the artificial chromosomes and to
10 further isolate, manipulate and introduce into other cells artificial chromosomes carrying unique genetic compositions.

Thus, the use of artificial chromosomes and homologous recombination in plant cells can be used to isolate and identify many valuable crop traits.

15 In addition to the use of artificial chromosomes for the isolation and testing of large regions of naturally occurring DNA, methods for the use of artificial chromosomes and cloned DNA are also contemplated. Similar to that described above, artificial chromosomes can be used to carry large regions of cloned DNA, including that derived from other plant species.

20 The ability to incorporate novel DNA elements into an artificial chromosome as it is being formed allows for the development of artificial chromosomes specifically engineered as a platform for testing of new genetic combinations, or "genomic" discoveries for model species such as *Arabidopsis*. It is known that specific "recombinase" systems can be
25 used in plant cells to excise or re-arrange genes. These same systems can be used to derive new gene combinations contained on an artificial chromosome.

The artificial chromosomes can be engineered as platforms to accept large regions of cloned DNA, such as that contained in Bacterial

-85-

Artificial Chromosomes (BACs) or Yeast Artificial Chromosomes (YACs). It is further contemplated, that as a result of the typical structure of artificial chromosomes containing tandemly repeated DNA blocks, that sequences other than cloned DNA sequence can be introduced by

5 recombination processes. In particular recombination within a predefined region of the tandemly repeated DNA within the artificial chromosome provides a mechanism to "stack" numerous regions of cloned DNA, including large regions of DNA contained within BACs or YACs clones. Thus, multiple combinations of genes can be introduced onto artificial

10 chromosomes and these combinations tested for functionality. In particular, it is contemplated that multiple YACs or BACs can be stacked onto an artificial chromosomes, the BACs or YACs containing multiple genes of complex pathways or multiple genetic pathways. The BACs or YACs are typically selected based on genetic information available within

15 the public domain, for example from the Arabidopsis Information Management System (<http://aims.cps.msu.edu/aims/index.html>) or the information related to the plant DNA sequences available from the Institute for Genomic Research (<http://www.tigr.org>) and other sites known to those skilled in the art. Alternatively, clones can be chosen at

20 random and evaluated for functionality. It is contemplated that combinations providing a desired phenotype can be identified by isolation of the artificial chromosome containing the combination and analyzing the nature of the inserted cloned DNA.

In this regard, it is contemplated that the use of site-specific

25 recombination sequences can have considerable utility in developing artificial chromosomes containing DNA sequences recognized by recombinase enzymes and capable of accepting DNA sequences containing same. The use of site-specific recombination as a means to target an introduced DNA to a specific locus has been demonstrated in

-86-

the art and such methods can be employed. The recombinase systems can also be used to transfer the cloned DNA regions contained within the artificial chromosome to the naturally occurring plant or mammalian chromosomes.

- 5 As noted herein, many site-specific recombinases are known and can be identified (Kilby *et al.* (1993) *Trends in Genetics* 9:413-418). The three recombinase systems that have been extensively employed include: an activity identified as R encoded by the pSR1 plasmid of *Zygosaccharomyces rouxii*, FLP encoded for the 2 μ m circular plasmid from
10 *Saccharomyces cerevisiae* and *Cre-lox* from the phage P1.

- The integration function of site-specific recombinases is contemplated as a means to assist in the derivation of genetic combinations on artificial chromosomes. In order to accomplish this, it is contemplated that a first step of introducing site-specific recombinase
15 sites into the genome of a plant cell in an essentially random manner is conducted, such that the plant cell has one or more site-specific recombinase recognition sequences on one or more of the plant chromosomes. An artificial chromosome is then introduced into the plant cell, the artificial chromosome engineered to contain a recombinase
20 recognition site (e.g., integration site) capable of being recognized by a site-specific recombinase. Optionally, a gene encoding a recombinase enzyme is also included, preferably under the control of an inducible promoter. Expression of the site-specific recombinase enzyme in the plant cell, either by induction of a inducible recombinase gene, or
25 transient expression of a recombinase sequence, causes a site-specific recombination event to take place, leading to the insertion of a region of the plant chromosomal DNA (containing the recombinase recognition site) into the recombinase recognition site of the artificial chromosome, and forming an artificial chromosome containing plant chromosomal DNA.

-87-

The artificial chromosome can be isolated and introduced into a heterologous host, preferably a plant host, and expression of the newly introduced plant chromosomal DNA can be monitored and evaluated for desirable phenotypic changes. Accordingly, carrying out this

5 recombination with a population of plant cells wherein the chromosomally located recombinase recognition site is randomly scattered throughout the chromosomes of the plant, can lead to the formation of a population of artificial chromosomes, each with a different region of plant chromosomal DNA, and each potentially representing a novel genetic combination.

10 This method requires the precise site-specific insertion of chromosomal DNA into the artificial chromosome. This precision has been demonstrated in the art. For example, Fukushige and Sauer ((1992) Proc. Natl. Acad. Sci. USA, 89:7905-7909) demonstrated that the *Cre-lox* homologous recombination system could be successfully employed to
15 introduce DNA into a predefined locus in a chromosome of mammalian cells. In this demonstration a promoter-less antibiotic resistance gene modified to include a *lox* sequence at the 5' end of the coding region was introduced into CHO cells. Cells were re-transformed by electroporation with a plasmid that contained a promoter with a *lox* sequence and a
20 transiently expressed *Cre* recombinase gene. Under the conditions employed, the expression of the *Cre* enzyme catalyzed the homologous recombination between the *lox* site in the chromosomally located promoter-less antibiotic resistance gene, and the *lox* site in the introduced promoter sequence, leading to the formation of a functional antibiotic
25 resistance gene. The authors demonstrated efficient and correct targeting of the introduced sequence, 54 of 56 lines analyzed corresponded to the predicted single copy insertion of the DNA due to *Cre* catalyzed site-specific homologous recombination between the *lox* sequences.

-88-

Accordingly a *lox* sequence may be first added to a genome of a plant species capable of being transformed and regenerated to a whole plant to serve as a recombinase target DNA sequence for recombination with an artificial chromosome. The *lox* sequence may be optimally
5 modified to further contain a selectable marker which is inactive but can be activated by insertion of the *lox* recombinase recognition sequence into the artificial chromosome.

A promoterless marker gene or selectable marker gene linked to the recombinase recognition sequence, which is first inserted into the
10 chromosomes of a plant cell can be used to engineer a platform chromosome. A promoter is linked to a recombinase recognition site, in an orientation that allows the promoter to control the expression of the marker or selectable marker gene upon recombination within the artificial chromosome. Upon a site-specific recombination event between a
15 recombinase recognition site in a plant chromosome and the recombinase recognition site within the introduced artificial chromosome, a cell is derived with a recombined artificial chromosome, the artificial chromosome containing an active marker or selectable marker activity that permits the identification and or selection of the cell.

20 The artificial chromosomes can be transferred to other plant or animal species and the functionality of the new combinations tested. The ability to conduct such an inter-chromosomal transfer of sequences has been demonstrated in the art. For example, the use of the *Cre-lox* recombinase system to cause a chromosome recombination event
25 between two chromatids of different chromosomes has been shown.

Any number of recombination systems may be employed as described herein, such as, but not limited to, bacterially derived systems such as the att/int system of phage lambda, and the Gin/gix system.

-89-

More than one recombination system may be employed, including, for example, one recombinase system for the introduction of DNA into an artificial chromosome, and a second recombinase system for the subsequent transfer of the newly introduced DNA contained within an
5 artificial chromosome into the naturally occurring chromosome of a second plant species. The choice of the specific recombination system used will be dependent on the nature of the modification contemplated.

By having the ability to isolate an artificial chromosome, in particular, artificial chromosomes containing plant chromosomal DNA
10 introduced via site-specific recombination, and re-introduce the chromosome into other mammalian or plant cells, particularly plant cells, these new combinations can be evaluated in different crop species without the need to first isolate and modify the genes, or carry out multiple transformations or gene transfers to achieve the same
15 combination isolation and testing combinations of the genes in plants. The use of a site-specific recombinase also allows the convenient recovery of the plant chromosomal region into other recombinant DNA vectors and systems, such as mammalian or insect systems, for manipulation and study.

20 Also contemplated herein are *ACes*, cell lines and methods for use in screening a new chromosomal combinations, deletions, truncations with eucaryotic genome that take advantage of the site-specific recombination systems incorporated onto platform *ACes* provided herein. For example, provided herein is a cell line useful for making a library of
25 *ACes*, comprising a multiplicity of heterologous recombination sites randomly integrated throughout the endogenous chromosomes. Also provided herein is a method of making a library of *ACes* comprising random portions of a genome, comprising introducing one or more *ACes* into a cell line comprising a multiplicity of heterologous recombination

-90-

sites randomly integrated throughout the endogenous chromosomes, under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites within the cell's chromosomal DNA; and isolating said

5 multiplicity of *ACes*, thereby producing a library of *ACes* whereby multiple *ACes* have different portions of the genome within. Also provided herein is a library of cells useful for genomic screening, said library comprising a multiplicity of cells, wherein each cell comprises an *ACes* having a mutually exclusive portion of a chromosomal nucleic acid therein. The

10 library of cells can be from a different species and/or cell type than the chromosomal nucleic acid within the *ACes*. Also provided is a method of making one or more cell lines, comprising

- a) integrating into endogenous chromosomal DNA of a selected cell species, a multiplicity of heterologous recombination sites,
- 15 b) introducing a multiplicity of *ACes* under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites integrated within the cell's endogenous chromosomal DNA;
- c) isolating said multiplicity of *ACes*, thereby producing a library of
- 20 *ACes* whereby a multiplicity of *ACes* have mutually exclusive portions of the endogenous chromosomal DNA therein;
- d) introducing the isolated multiplicity of *ACes* of step c) into a multiplicity of cells, thereby creating a library of cells;
- e) selecting different cells having mutually exclusive *ACes* therein
- 25 and clonally expanding or differentiating said different cells into clonal cell cultures, thereby creating one or more cell lines.

These *ACes*, cell lines and methods utilize the site-specific recombination sites on platform *ACes* analogous YAC manipulation related to: the methods of generating terminal deletions in normal and

-91-

artificial chromosomes (e.g., *ACes*; as described in Vollrath et al., 1988, *PNAS, USA*, 85:6027-66031; and Pavan et al., *PNAS, USA*, 87:1300-1304); the methods of generating interstitial deletions in normal and artificial chromosomes (as described in Campbell et al., 1991, *PNAS, USA*, 888:5744-5748); and the methods of detecting homologous recombination between two *ACes* (as described in Cellini et al., 1991, *Nuc. Acid Res.*, 19(5):997-1000).

5. Use of platform *ACes* in Pharmacogenomic/toxicology applications (development of "Reporter *ACes*")

In addition to the placement of genes onto *ACes* chromosomes for therapeutic protein production or gene therapy, the platform can be engineered via the IntR lambda integrase to carry reporter-linked constructs (reporter genes) that monitor changes in cellular physiology as measured by the particular reporter gene (or a series of different reporter genes) readout. The reporter linked constructs are designed to include a gene that can be detected (by for example fluorescence, drug resistance, immunohistochemistry, or transcript production, and the like) with well-known regulatory sequences that would control the expression of the detectable gene. Exemplary regulatory promoter sequences are well-known in the art:

A) Reporter *ACes* for drug pathway screening

The *ACes* can be engineered to carry reporter-linked constructs that indicate a signal is being transduced through one or a number of pathways. For example, transcriptionally regulated promoters from genes at the end (or any other chosen point) of particular signal transduction pathways could be engineered on the *ACes* to express the appropriate readout (either by fluorescent protein production or drug resistance) when the pathway is activated (or down-regulated as well). In one embodiment, a number of reporters from different can be placed on a

-92-

ACes chromosome. Cells (and/or whole animals) containing such a Reporter *ACes* could be exposed to a variety of drugs or compounds and monitored for the effects of the drugs or compounds upon the selected pathway(s) by the reporter gene(s). Thus, drugs or compounds can be
5 classified or identified by particular pathways they excite or down-regulate. Similarly, transcriptional profiles obtained from genomic array experiments can be biologically validated using the reporter *ACes* provided herein.

B) Reporter *ACes* for toxic compound testing

10 Environmental or man-made genotoxicants can be tested in cell lines carrying a number of reporter-genes platform *ACes* linked to promoters that are transcriptionally regulated in response to DNA damage, induced apoptosis or necrosis, and cell-cycle perturbations. Furthermore, new drugs and/or compounds could be tested in a similar manner with the
15 genotoxicant *ACes* reporter for their cellular/genetic toxicity by such a screen. Likewise, toxic compound testing could be carried out in whole transgenic animals carrying the *ACes* chromosome that measures genotoxicant exposure ("canary in a coal mine"). Thus, the same or similar type *ACes* could be used for toxicity testing in either a cell-based
20 or whole animal setting. An example would include *ACes* that carry reporter-linked genes controlled by various cytochrome P450 profiled promoters and the like.

C) Reporter *ACes* for individualized pharmacogenomics/drug profiling

25 A common disease may arise via various mechanisms. In many instances there are multiple treatments available for a given disease. However, the success of a given treatment may depend upon the mechanism by which the disease originated and/or by the genetic background of the patient. In order to establish the most effective

-93-

treatment for a given patient one could utilize the *ACes* reporters provided herein. *ACes* reporters can be used in patient cell samples to determine an individualized drug regimen for the patient. In addition, potential polymorphisms affecting the transcriptional regulation of an individual's particular gene can be assessed by this approach.

D) Reporter *ACes* for classification of similar patient tumors

As with other diseases as described in 5.C) above, cancer cells arise via different mechanisms. Furthermore, as a cancerous cell propagates it may undergo genomic alterations. An *ACes* reporter transferred to cells of different patients having the same disease, i.e. similar cancers, could be used to categorize the particular cancer of each patient, thereby facilitating the identification of the most effective therapeutic regimen. Examples would include the validation of array profiling of certain classes of breast cancers. Subsequently, appropriate drug profiling could be carried out as described above.

E) Reporter *ACes* as a "differentiation" sensor

Using the *ACes* reporter as a "differentiation" sensor in stem cells or other progenitor cells in order to enrich by selection (either FACS based screening, drug selection and/or use of suicide gene) for a particular class of differentiated or undifferentiated cells. For example, in one embodiment, this assay could also be used for compound screening for small molecule modifiers of cell differentiation.

F) Whole animal studies with Reporter *ACes*

Finally, with whole-body fluorescence imaging technology (Yang et al. (2000) PNAS 97:12278) any of the above Reporter *ACes* methods could be used in conjunction with whole-body imaging to monitor reporter genes within whole animals without sacrificing the animal. This would allow temporal and spatial analysis of expression patterns under a given set of conditions. The conditions tested may include for example, normal

-94-

differentiation of a stem cell, response to drug or compound treatment whether targeted to the diseased tissue or presented systemically, response to genotoxicants, and the like.

The following examples are included for illustrative purposes only
5 and are not intended to limit the scope of the invention.

EXAMPLE 1

pFK161

Cosmid pFK161 (SEQ ID NO: 118) was obtained from Dr. Gyula Hadlaczký and contains a 9 kb *NotI* insert derived from a murine rDNA
10 repeat (see clone 161 described in PCT Application Publication No. WO97/40183 by Hadlaczký *et al.* for a description of this cosmid). This cosmid, referred to as clone 161 contains sequence corresponding to nucleotides 10,232-15,000 in SEQ ID NO. 26. It was produced by inserting fragments of the megachromosome (see, U.S. Patent No.
15 6,077,697 and International PCT application No. WO 97/40183). For example, H1D3, which was deposited at the European Collection of Animal Cell Culture (ECACC) under Accession No. 96040929, is a mouse-hamster hybrid cell line carrying this megachromosome into plasmid pWE15 (Stratagene, La Jolla, California; SEQ ID No. 31) as
20 follows. Half of a 100 μ l low melting point agarose block (mega-plug) containing isolated SATACs was digested with *NotI* overnight at 37°C. Plasmid pWE15 was similarly digested with *NotI* overnight. The mega-plug was then melted and mixed with the digested plasmid, ligation buffer and T4 DNA ligase. Ligation was conducted at 16°C overnight. Bacterial
25 DH5 α cells were transformed with the ligation product and transformed cells were plated onto LB/Amp plates. Fifteen to twenty colonies were grown on each plate for a total of 189 colonies. Plasmid DNA was isolated from colonies that survived growth on LB/Amp medium and analyzed by Southern blot hybridization for the presence of DNA that

-95-

hybridized to a pUC19 probe. This screening methodology assured that all clones, even clones lacking an insert but yet containing the pWE15 plasmid, would be detected.

Liquid cultures of all 189 transformants were used to generate
5 cosmid minipreps for analysis of restriction sites within the insert DNA. Six of the original 189 cosmid clones contained an insert. These clones were designated as follows: 28 (~9-kb insert), 30 (~9-kb insert), 60 (~4-kb insert), 113 (~9-kb insert), 157 (~9-kb insert) and 161 (~9-kb insert). Restriction enzyme analysis indicated that three of the clones
10 (113, 157 and 161) contained the same insert.

For sequence analysis the insert of cosmid clone no. 161 was subcloned as follows. To obtain the end fragments of the insert of clone no. 161, the clone was digested with *NotI* and *BamHI* and ligated with *NotI/BamHI*-digested pBluescript KS (Stratagene, La Jolla, California).
15 Two fragments of the insert of clone no. 161 were obtained: a 0.2-kb and a 0.7-kb insert fragment. To subclone the internal fragment of the insert of clone no. 161, the same digest was ligated with *BamHI*-digested pUC19. Three fragments of the insert of clone no. 161 were obtained: a 0.6-kb, a 1.8-kb and a 4.8-kb insert fragment.

20 The insert corresponds to an internal section of the mouse ribosomal RNA gene (rDNA) repeat unit between positions 7551-15670 as set forth in GENBANK accession no. X82564, which is provided as SEQ ID NO. 18. The sequence data obtained for the insert of clone no. 161 is set forth in SEQ ID NOS. 19-25. Specifically, the individual
25 subclones corresponded to the following positions in GENBANK accession no. X82564 (SEQ ID NO:18) and in SEQ ID NOs. 19-25:

-96-

5

Subclone	Start	End	Site	SEQ ID No.
	in X82564			
161k1	7579	7755	<i>NotI</i> , <i>Bam</i> HI	19
161m5	7756	8494	<i>Bam</i> HI	20
161m7	8495	10231	<i>Bam</i> HI	21 (shows only sequence corresponding to nt. 8495-8950), 22 (shows only sequence corresponding to nt. 9851- 10231)
161m12	10232	15000	<i>Bam</i> HI	23 (shows only sequence corresponding to nt. 10232-10600), 24 (shows only sequence corresponding to nt. 14267-15000)
161k2	15001	15676	<i>NotI</i> , <i>Bam</i> HI	25

10 The sequence set forth in SEQ ID NOs. 19-25 diverges in some positions from the sequence presented in positions 7551-15670 of GENBANK accession no. X82564. Such divergence may be attributable to random mutations between repeat units of rDNA.

15 For use herein, the rDNA insert from the clone was prepared by digesting the cosmid with *NotI* and *Bgl*II and was purified as described above. Growth and maintenance of bacterial stocks and purification of plasmids were performed using standard well known methods (see, *e.g.*, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press), and plasmids were purified from bacterial cultures using Midi - and Maxi-preps Kits (Qiagen,
20 Mississauga, Ontario).

pDsRed1N1

25 This vector is available from Clontech (see SEQ ID No. 29) and encodes the red fluorescent protein (DsRed; Genbank accession no. AF272711; SEQ ID Nos. 39 and 40). DsRed, which has a vivid red fluorescence, was isolated from the IndoPacific sea anemone relative *Discosoma* species. The plasmid pDsRed1N1 (Clontech; SEQ ID No. 29) constitutively expresses a human codon-optimized variant of the

-97-

fluorescent protein under control of the CMV promoter. Unmodified, this vector expresses high levels of DsRed1 and includes sites for creating N-terminal fusions by cloning proteins of interest into the multiple cloning site (MCS). It is Kan and Neo resistant for selection in bacterial or
5 eukaryotic cells.

Plasmid pMG

Plasmid pMG (InvivoGen, San Diego, California; see SEQ. ID. NO. 27 for the nucleotide sequence of pMG) contains the hygromycin phosphotransferase gene under the control of the immediate-early human
10 cytomegalovirus (hCMV) enhancer/promoter with intron A. Vector pMG also contains two transcriptional units allowing for the coexpression of two heterologous genes from a single vector sequence.

The first transcriptional unit of pMG contains a multiple cloning site for insertion of a gene of interest, the hygromycin phosphotransferase
15 gene (*hph*) and the immediate-early human cytomegalovirus (hCMV) enhancer/promoter with intron A (see, *e.g.*, Chapman *et al.* (1991) *Nuc. Acids Res.* 19:3979-3986) located upstream of *hph* and the multiple cloning site, which drives the expression of *hph* and any gene of interest inserted into the multiple cloning site as a polycistronic mRNA. The first
20 transcriptional unit also contains a modified EMCV internal ribosomal entry site (IRES) upstream of the *hph* gene but downstream of the hCMV promoter and MCS for ribosomal entry in translation of the *hph* gene (see SEQ ID NO. 27, nucleotides 2736-3308). The IRES is modified by insertion of the constitutive *E. coli* promoter (EM7) within an intron (IM7)
25 into the end of the IRES. In mammalian cells, the *E. coli* promoter is treated as an intron and is spliced out of the transcript. A polyadenylation signal from the bovine growth hormone (bGh) gene (see, *e.g.*, Goodwin and Rottman (1992) *J. Biol. Chem.* 267:16330-16334) and a pause site derived from the 3' flanking region of the human $\alpha 2$

-98-

globin gene (see, *e.g.*, Enriquez-Harris *et al.* (1991) *EMBO J.* 10:1833-1842) are located at the end of the first transcription unit. Efficient polyadenylation is facilitated by inserting the flanking sequence of the bGh gene 3' to the standard AAUAAA hexanucleotide sequence.

5 The second transcriptional unit of pMG contains another multiple cloning site for insertion of a gene of interest and an EF-1 α /HTLV hybrid promoter located upstream of this multiple cloning site, which drives the expression of any gene of interest inserted into the multiple cloning site. The hybrid promoter is a modified human elongation factor-1 alpha (EF-1
10 alpha) gene promoter (see, *e.g.*, Kim *et al.* (1990) *Gene* 91:217-223) that includes the R segment and part of the U5 sequence (R-U5') of the human T-cell leukemia virus (HTLV) type I long terminal repeat (see, *e.g.*, Takebe *et al.* (1988) *Mol. Cell. Biol.* 8:466-472). The Simian Virus 40 (SV40) late polyadenylation signal (see Carswell and Alwine (1989) *Mol.*
15 *Cell. Biol.* 9:4248-4258) is located downstream of the multiple cloning site. Vector pMG contains a synthetic polyadenylation site for the first and second transcriptional units at the end of the transcriptional unit based on the rabbit β -globin gene and containing the AATAAA hexanucleotide sequence and a GT/T-rich sequence with 22-23
20 nucleotides between them (see, *e.g.*, Levitt *et al.* (1989) *Genes Dev.* 3:1019-1025). A pause site derived from the C2 complement gene (see, Moreira *et al.* (1995) *EMBO J.* 14:3809-3819) is also located at the 3' end of the second transcriptional unit.

Vector pMG also contains an ori sequence (ori pMB1) located
25 between the SV40 polyadenylation signal and the synthetic polyadenylation site.

EXAMPLE 2

A. Construction of targeting vector and transfection into LMtk- cells for the generation of platform chromosomes

-99-

A targeting vector derived from the vector pWE15 (GeneBank Accession # X65279) was modified by replacing the *SaI* (Klenow filled)/*SmaI* neomycin resistance containing fragment with the *PvuII*/*Bam*HI (Klenow filled) puromycin resistance containing fragment (isolated from plasmid pPUR, Clontech Laboratories, Inc. Palo Alto, CA; SEQ ID No. 30) resulting in plasmid pWEPuro. Subsequently a 9 Kb *NotI* fragment from the plasmid pFK161 (SEQ ID NO: 118) containing a portion of the mouse rDNA region was cloned into the *NotI* site of pWEPuro resulting in plasmid pWEPuro9K (Figure 2). The vector pWEPuro9K was digested with *SpeI* to linearize and transfected into LMtk- mouse cells. Puromycin resistant colonies were isolated and subsequently tested for artificial chromosome formation via fluorescent *in situ* hybridization (FISH) (using mouse major and minor DNA repeat sequences, the puromycin gene and telomeres sequences as probes), and fluorescent activated cell sorting (FACS). From this sort, a subclone was isolated containing an artificial chromosome, designated 5B11.12, which carries 4-8 copies of the puromycin resistance gene contained on the pWEPuro9K vector. FISH analysis of the 5B11.12 subclone demonstrated the presence of telomeres and mouse minor on the ACes. DOT PCR has been done on the 5B11.12 ACes revealing the absence of uncharacterized euchromatic regions on the ACes. A recombination site, such as an *att* or *loxP* engineering site or a plurality thereof, was introduced onto this ACes thereby providing a platform for site-specific introduction of heterologous nucleic acid.

B. Targeting a single sequence specific recombination site onto platform chromosomes

After the generation of the 5B11.12 platform, a single sequence-specific recombination site is placed onto the platform chromosome via homologous recombination. For this, DNA sequences containing the site-

-100-

specific recombination sequence can be flanked with DNA sequences of homology to the platform chromosome. For example, using the platform chromosome made from the pWEPuro9K vector, mouse rDNA sequences or mouse major satellite DNA can be used as homologous sequences to
5 target onto the platform chromosome. A vector is designed to have these homologous sequences flanking the site-specific recombination site and, after the appropriate restriction enzyme digest to generate free ends of homology to the platform chromosome, the DNA is transfected into cells harboring the platform chromosome (Figure 3). Examples of site-specific
10 cassettes that are targeted to the platform chromosome using either mouse rDNA or mouse major repeat DNA include the SV40-attP-hygro cassette and a red fluorescent protein (RFP) gene flanked by loxP sites (Cre/lox, see, *e.g.*, U.S. Patent No. 4,959,317 and description herein). After transfection and integration of the site-specific cassette,
15 homologous recombination events onto the platform chromosome are subcloned and identified by FACS (e.g. screen and single cell subclone via expression of resistance or fluorescent marker) and PCR analysis.

For example, a vector can be constructed containing regions of the mouse rDNA locus flanking a gene cassette containing the SV40 early
20 reporter-bacteriophage lambda attP site-hygromycin selectable marker (see Figure 4 and described below). The use of the bacteriophage lambda attP site for lambda integrase-mediated site-specific recombination is described below. Homologous recombination event of the SV40-attP-hygro cassette onto the platform chromosome was identified using PCR
25 primers that detect the homologous recombination and further confirmed by FISH analysis. After identifying subcloned colonies containing the platform chromosome with a single site-specific recombination site, cells carrying the platform chromosome with a single site-specific

-101-

recombination site can now be engineered with site-specific recombinases (e.g. lambda INT, Cre) for integrating a target gene expression vector.

**C. Targeting a red fluorescent protein (RFP) gene flanked by loxP sites
5 onto 5B11.12 platform**

As another example, while loxP recombination sites could have been introduced onto the *ACes* during *de novo* biosynthesis, it was thought that this might result in multiple segments of the *ACes* containing a high number of loxP sites, potentially leading to instability upon Cre-
10 mediated recombination. A gene targeting approach was therefore devised to introduce a more limited number of loxP recombination sites into a locus of the 5B11-12 *ACes* containing introduced and possibly co-amplified endogenous rDNA sequences. Although there are more than 200 copies of rDNA genes in the haploid mouse genome distributed
15 amongst 5-11 chromosomes (depending on strain), rDNA sequences were chosen as the target on the *ACes* since they represent a less frequent target than that of the satellite repeat sequences. Moreover, having observed much stronger pWEPuro9K hybridization to the 5B11-12 *ACes* than to other LMTK⁻ chromosomes and in light of the observation that the
20 transcribed spacer sequences within the rDNA may be less conserved than the rRNA coding regions, it was contemplated that a targeting vector based on the rDNA gene segment in pWEPuro9K would have a higher probability of targeting to the *ACes* rather than to other LMTK⁻ chromosomes. Accordingly, a targeting vector, pBSFKLoxDsRedLox, was
25 designed and constructed based on the rDNA sequences contained in pWEPuro9K.

The plasmid pBSFKLoxDsRedLox was generated in 4 steps. First, the *NotI* rDNA insert of pWEPuro9K (Figure 2) was inserted into pBS SK- (Stratagene) giving rise to pBSFK. Second, a loxP polylinker cassette was

-102-

generated by PCR amplification of pNEB193 (SEQ ID NO:32; New England Biolabs) using primers complementary to the M13 forward and reverse priming sites at their 3' end and a 34 bp 5' extension comprising a LoxP site. This cassette was reinserted into pNEB193 generating
5 p193LoxMCSLox. Third, the DsRed gene from pDsRed1-N1 (SEQ ID NO:29; Clontech) was then cloned into the polylinker between the loxP sites generating p193LoxDsRedLox. Fourth, a fragment consisting of the DsRed gene flanked by loxP sites was cloned into a unique *NdeI* within the rDNA insert of pBSFK generating pBSFKLoxDsRedLox.

10 A gel purified 11 Kb *PmlI* /*EcoRV* fragment of pBSFKLoxDsRedLox was used for transfection. To detect targeted integration, PCR primers were designed from rDNA sequences within the 5' *NotI*-*PmlI* fragment of pWEPuro9K that is not present on the targeting fragment (5' primer) and sequence within the LoxDsRedLox cassette (3' primer). If the targeting
15 DNA integrated correctly within the rDNA sequences, PCR amplification using these primers would give rise to a 2.3 Kb band. PCR reactions containing 1-4 μ l of genomic DNA were carried out according to the MasterTaq protocol (Eppendorf), using murine rDNA 5' primer (5'-CGGACAATGCGGTTGTGCGT-3'; SEQ ID NO:72) and DsRed 3' primer
20 (5'-GGCCCCGTAATGCAGAAGAA-3'; SEQ ID NO:73) and PCR products were analyzed by agarose gel electrophoresis.

1.5X10⁶ 5B11-12 LMTK⁻ cells were transfected with 2 μ g of the pBSFKLoxDsRedLox targeting DNA described above using Lipofectamine Plus (Invitrogen). For flow sorting, harvested cells were suspended in
25 medium and applied to the Becton Dickinson Vantage SE cell sorter, equipped with 488 nm lasers for excitation and 585/42 bandpass filter for optimum detection of RFP fluorescence. Cells were sorted using dPBS as sheath buffer. Negative control parental 5B11-12 cells and a positive control LMTK⁻ cell line stably transfected with DsRed were used to

-103-

establish the selection gates. The RFP positive gated populations were recovered, diluted in medium supplemented with 1X penicillin-streptomycin (Invitrogen), then plated and cultured as previously described. After 4 rounds of enrichment, the percentage of RFP positive
5 cells reached levels of 50% or higher. DNA from populations was analyzed by PCR for evidence of targeted integration. Ultimately, single cell subclones were established from positive pools and were analyzed by PCR and PCR-positive clones confirmed by FISH as described below. DNA was purified from pools or single cell clones using previously
10 described methods set forth in Lahm et al., Transgenic Res., 1998; 7:131-134, or in some cases using a Wizard Genomic DNA purification kit (Promega). For FISH analysis, a biotinylated DsRed gene probe was generated by PCR using DsRed specific primers and biotin-labeled dUTP (5' RFP primer: 5'-GGTTTAAAGTGCCTCCTCCAAGAACGTCATC-3',
15 SEQ ID NO:74; and 3' RFP primer: 5'-AGATCTAGAGCCGCGCTACAGGAACAGGTGGTGGCGGCC-3'; SEQ ID NO:75). To maximize the signal intensity of the DsRed probe, Tyramide amplification was carried out according to the manufacturers protocols (NEN).

20 The process of testing the feasibility of a more general targeting strategy that would not rely on enrichment *via* drug selection of stably transfected clones can be summarized as follows. A red fluorescent protein gene (RFP; encoded by the DsRed gene) was inserted between the loxP sites of the targeting vector to form pBSFKLoxDsRedLox. After
25 transfection with pBSFKLoxDsRedLox, sequential rounds of high speed flow sorting and expansion of sorted cells in culture could then be used to enrich for stable transformants expressing RFP. In the event of targeted integration, PCR screening with primers that amplify from a spacer region within the segment of the 45s pre-rRNA gene in pWEPuro9K to a specific

-104-

anchor sequence within the DsRed gene in the targeting cassette would give rise to a diagnostic 2.3 Kb band. However, as rDNA clusters are found on several chromosomes, confirmation of targeting to an *ACes* would require fluorescence in situ hybridization (FISH) analysis. Finally, the flanking of the DsRed gene by loxP sites would allow for its removal and subsequent replacement with other genes of interest.

After transfection of the targeting sequence into 5B11-12 cells, enrichment for targeted clones was carried out using a combination of flow cytometry to detect red-fluorescing cells and PCR screening.

10 Ultimately 17 single cell subclones were identified as potential targeted clones by PCR and of these 16 were found by FISH to contain the DsRed integration event into the *ACes*. These subclones are referred to herein as D11-C4, D11-C12, D11-H3, C9-C9, C9-B9, C9-F4, C9-H8, C9-F2, C9-G8, C9-B6, C9-G3, C9-E12, C9-A11, C11-E3, C11-A9 and C11-H4. PCR
15 analysis of genomic DNA isolated from the D11-C4 subclone gave rise to a 2.3 Kb band, indicative of a targeted integration into an rDNA locus. Further analysis of the subclone by FISH analysis with a DsRed gene probe demonstrated integration of the LoxDsRedLox targeting cassette on the *ACes* co-localizing with one of the regions of rDNA staining seen on
20 the 5B11-12 *ACes*, consistent with a targeted integration into an rDNA locus of the *ACes*, while integrations on other chromosomes were not observed. Since transfected cells were maintained as heterogeneous populations through several cycles of sorting and replating it was not possible to estimate the frequency of targeted events. In most
25 mammalian cell lines the frequency of gene targeting via homologous recombination is roughly 10^{-5} - 10^{-7} treated cells. Despite the low frequency of these events in mammalian cells, it is clear that an RFP expression based screening paradigm, coupled with PCR analysis, can effectively detect and enrich for such infrequent events in a large

-105-

population. In instances where drug selection is not possible or not desirable, such a system may provide a useful alternative. It was also verified that the modified *ACes* in subclone D11-C4 could be purified by flow cytometry. The results indicate that the flow karyogram of the D11-C4 subclone was unaltered from that of the 5B11-12 cell line. Thus, the D11-C4 *ACes* can be purified in high yield from native chromosomes of the host cell line.

D. Reduction of LoxP on *ACes* to a single site.

10 The strong hybridization signal detected by FISH on the *ACes* using the DsRed gene probe suggests that several copies of the targeting cassette may be present on the *ACes* in the D11-C4 line. This also suggests that multiple rDNA genes have been correctly targeted.

15 Accordingly, in certain embodiments where necessary, the number of loxP sites on the *ACes* can be reduced to a single site by *in situ* treatment with *Cre* recombinase, provided that the sites are co-linear. Such a process is described for multiple loxP-flanked integrations on a native mouse chromosome (Garrick et al., Nature Genet., 1998, Jan;18(1):56-59). Reduction to a single loxP site on the D11-C4 *ACes* would result in the loss of the DsRed gene, forming the basis of a useful screen for this event.

20 For this purpose, a Cre expression plasmid pCX-Cre/GFP III has been generated by first deleting the EcoRI fragment of pCX-eGFP (SEQ ID NO:71) containing the eGFP coding sequence and replacing it with that of a PCR amplified Cre recombinase coding sequence (SEQ ID NO:58), generating pCX-Cre. Next, the *Asel*/*Sspl* fragment of pD2eGFP-N1 (containing the CMV promoter driving the D2EGFP gene with SV40 polyA signal; Clontech; SEQ ID NO:87) was inserted into the filled *HindIII* site of pCX-Cre, generating pCX-Cre\GFP III. Control plasmid pCX-CreRev\GFP

-106-

III was generated in similar fashion except that the Cre recombinase coding sequence was inserted in the antisense orientation. LMTK⁻ cell line D11-C4 (containing first generation platform ACes with multiple loxP-DsRED sites) and 5B11-12 cell line (containing ACes with no loxP-DsRED sites) are maintained in culture as described above. D11C4 cells are transfected with 2 μ g of plasmid pCX-Cre\GFP III or 2 μ g pCX-CreRev\GFP III using Lipofectamine (Invitrogen) as previously described.

Forty-eight to seventy-two hours after transfection, transfected D11-C4 cells are harvested and GFP positive cells are sorted by cell cytometry using a FACSta Vantage cell sorter (Beckton-Dickinson) as follows: All D11-C4 cells transfected with pCX-Cre\GFP III or control plasmid pCX-CreRev\GFP III that exhibit GFP fluorescent higher than the gate level established by untransfected cells are collected and placed in culture a further 7-14 days. After 7-14 days the initial D11-C4 cells are harvested and analyzed by cell cytometry as follows: Untransfected D11-C4 cells are used to establish the gate that defines the RFP positive population, while 5B11-12 cells are used to set the RFP negative gate. The GFP positive population of D11-C4 transfected with pCX-Cre\GFP III should show decreased red fluorescence compared to pCX-CreRev\GFP III transfected or untransfected control D11-C4 cells. The cells exhibiting greatly decreased or no RFP expression are collected and single cell clones subsequently established. These clones will be expanded and analyzed by fluorescence *in-situ* hybridization and Southern blotting to confirm the removal of loxP-DsRed gene copies.

25

EXAMPLE 3

Construction of targeting vector and transfection into LMtk⁻ cells for the generation of platform chromosomes containing multiple site-specific recombination sites

-107-

An example of a selectable marker system for the creation of a chromosome-based platform is shown in Figure 4. This system includes a vector containing the SV40 early promoter immediately followed by (1) a 282 base pair (bp) sequence containing the bacteriophage lambda attP site and (2) the puromycin resistance marker. Initially a *PvuII/StuI* fragment containing the SV40 early promoter from plasmid pPUR (Clontech Laboratories, Inc., Palo Alto, CA; Seq ID No. 30) was subcloned into the *EcoRI/CRI* site of pNEB193 (a PUC19 derivative obtained from New England Biolabs, Beverly, MA; SEQ ID No. 32) generating the plasmid pSV40193. The only differences between pUC19 and pNEB193 are in the polylinker region. A unique *AscI* site (GGCGCGCC) is located between the *BamHI* site and the *SmaI* site, a unique *PacI* site (TTAATTAA) is located between the *BamHI* site and the *XbaI* site and a unique *PmeI* site (GTTTAAAC) is located between the *PstI* site and the *SaI* site.

The attP site was PCR amplified from lambda genome (GenBank Accession # NC 001416) using the following primers:

attPUP: CCTTGCGCTAATGCTCTGTTACAGG SEQ ID No. 1

attPDWN: CAGAGGCAGGGAGTGGGACAAAATTG SEQ ID No. 2

After amplification and purification of the resulting fragment, the attP site was cloned into the *SmaI* site of pSV40193 and the orientation of the attP site was determined by DNA sequence analysis (plasmid pSV40193attP). The gene encoding puromycin resistance (Puro) was isolated by digesting the plasmid pPUR (Clontech Laboratories, Inc. Palo Alto, CA) with *AgeI/BamHI* followed by filling in the overhangs with Klenow and subsequently cloned into the *AscI* site downstream of the attP site of pSV40193attP generating the plasmid pSV40193attPsensePUR (Figure 4; SEQ ID NO:113)).

-108-

The plasmid pSV40193attPsensePUR was digested with *ScaI* and co-transfected with the plasmid pFK161 (SEQ ID NO: 118) into mouse LMtk- cells and platform artificial chromosomes were identified and isolated as described above. The process for generating this exemplary platform ACes containing multiple site-specific recombination sites is summarized in Figure 5. One platform ACes resulting from this experiment is designated B19-18. This platform ACes chromosome may subsequently be engineered to contain target gene expression nucleic acids using the lambda integrase mediated site-specific recombination system as described herein in Example 7 and 8.

EXAMPLE 4

Lambda integrase mediated site-specific recombination of a RFP expressing vector onto artificial chromosomes

In this example, a vector expressing the red fluorescent protein (RFP) was produced and recombined into the attP site residing on an artificial chromosome within LMTK- cells. This recombination is depicted in Figure 7.

A. Construction of expression vectors containing wildtype and mutant lambda integrase

Mutations at the glutamic acid at position 174 in the lambda integrase protein relaxes the requirement for the accessory protein IHF during recombination and DNA supercoiling *in vitro* (see, Miller *et al.* (1980) *Cell* 20:721-729; Lange-Gustafson *et al.* (1984) *J. Biol. Chem.* 259:12724-12732). Mutations at this site promote attP, attB intramolecular recombination in mammalian cells (Lorbach *et al.* (2000) *J. Mol. Biol* 296:1175-1181).

To construct nucleic acid encoding the mutant, lambda integrase was PCR amplified from bacteriophage lambda DNA (cl857 *ind Sam* 7; New England Biolabs) using the following primers:

Lamint1 (SEQ ID No. 3)

-109-

TTCGAATTCATGGGAAGAAGGCGAAGTCATGAGCG)

Lamint2 (SEQ ID No. 4)

(TTCGAATTCTTATTTGATTTCAATTTGTCCCAC).

- The resulting PCR product was digested with *EcoR* I and cloned into the
- 5 *EcoR* I site of pUC19. Lambda integrase was mutated at amino acid position 174 using QuikChange Site-Directed Mutagenesis Kit (Stratagene) and the following oligos (generating a glutamic acid to arginine change at position 174):

LambdaINTE174R

- 10 (SEQ ID No. 6)

(CGCGCAGCAAAATCTAGAGTAAGGAGATCAAGACTTACGGCTGACG),

LamintR174rev (SEQ ID No. 7)

(CGTCAGCCGTAAGTCTTGATCTCCTTACTCTAGATTTTGCTGCGCG).

The resulting site directed mutant was confirmed by sequence analysis.

- 15 The wildtype and mutant lambda genes were cloned into the *EcoR* I site of pCX creating pCX-LamInt (SEQ ID NO: 127) and pCXLamIntR (Figure 8; SEQ ID NO: 112).

- The plasmid pCX (SEQ ID No. 70) was derived from plasmid pCXeGFP (SEQ ID No. 71). Excision of the *EcoRI* fragment containing the
- 20 eGFP marker generated pCX. To generate plasmid pCXLamINTR (SEQ ID NO: 112) an *EcoRI* fragment containing the lambda integrase E174R (SEQ ID No. 37) mutation was cloned into the *EcoRI* site of pCX, and to generate plasmid pCX-LamINT, an *EcoRI* fragment containing the wild-type lambda integrase was cloned into the *EcoRI* site of pCX.

- 25 **B. Construction of integration vector containing attB and DsRed**

The plasmid pDsRedN1 (Clontech Laboratories, Palo Alto, CA; SEQ ID No. 29) was digested with *Hpa* I and ligated to the following annealed oligos:

attB1 (SEQ ID No. 8)

-110-

(TGAAGCCTGCTTTTTTATACTAACTTGAGCGAA)

attB2 (SEQ ID No. 9)

(TTCGCTCAAGTTAGTATAAAAAAGCAGGCTTCA)

The resulting vector (pDsRedN1-attB) was confirmed by PCR and
5 sequence analysis.

C. Transfection into LMtk- cells

LM(tk-) cells containing the Prototype A ACes (L1-18; Chromos
Molecular Systems Inc., Burnaby, BC Canada) were co-transfected with
pDsRedN1 or pDsRedN1-attB and either pCXLamInt (SEQ ID NO: 127) or
10 pCXLamIntR (SEQ ID NO: 112) using Lipofectamine Plus Reagent
(LifeTechnologies, Gaithersburg, MD). The transfected cells were grown
in DMEM (LifeTechnologies, Gaithersburg, MD) with 10% FBS (CanSera)
and G418 (CalBiochem) at a concentration of 1 mg/ml.

D. Enrichment by cell sorting

15 The transfected cells were sorted using a FACs Vantage SE cell
sorter (Becton Dickenson) to enrich for cells expressing DsRed. The cells
were excited with a 488 nm Argon laser at 200 watts and cells
fluorescing in the 585/42 detection channel were collected. The sorted
cells were returned to growth medium for recovery and expansion. After
20 three successive enrichments for cells expressing DsRed, single cell
sorting into 96 well plates was performed using the same parameters.
Duplicate plates of the single cell clones were made for PCR analysis.

E. PCR analysis of single cell clones

Pools of cells from each row and column of the 96 well plate were
25 used for DNA isolation. DNA was prepared using a Wizard Genomic DNA
purification kit (Promega Inc, Madison, WI). Nested PCR analysis on the
DNA pools was performed to confirm the site-specific recombination
event using the following primer sets:

-111-

attPdwn2 (SEQ ID No. 10)
(TCTTCTCGGGCATAAGTCGGACACC)

CMVen (SEQ ID No. 11)
(CTCACGGGGATTTCCTCAAGTCTCCAC)

5 followed by:

attPdwn (SEQ ID No. 12)
(CAGAGGCAGGGAGTGGGACAAAATTG)

CMVen2 (SEQ ID No. 13)
(CAACTCCGCCCCATTGACGCAAATG).

- 10 The resulting PCR reactions were analyzed by gel electrophoresis and the potential individual clones containing the site-specific recombination event were identified by combining the PCR results of all of the pooled rows and columns for each 96 well plate. The individual clones were then further analyzed by PCR using the following primers that flank the
- 15 recombination junction. L1for and F1rev flank the attR junction whereas REDfor and L2rev flank the attL junction (see Figure 7):

L1for (SEQ ID No. 14)
AGTATCGCCGAACGATTAGCTCTTCA

F1rev (SEQ ID No. 15)
20 GCCGATTTCGGCCTATTGGTTAAA

REDfor (SEQ ID No. 16)
CCGCCGACATCCCCGACTACAAGAA

L2rev (SEQ ID No. 17)
TTCCTTCGAAGGGGATCCGCCTACC.

25 **F. Sequence analysis of recombination junctions**

PCR products spanning the recombination junction were Topo-cloned into pcDNA3.1D/V5His (Invitrogen Inc., San Diego, CA) and then sequenced by cycle-sequencing. The clones were confirmed to have the correct *attR* and *attL* junctions by cycle sequencing.

30 **G. Fluorescent In Situ Hybridization (FISH)**

The cell lines containing the correct recombination junction sequence were further analyzed by fluorescent *in situ* hybridization (FISH)

-112-

by probing with the DsRed coding region labeled with biotin and visualizing with the Tyramide Signal Amplification system (TSA; NEN Life Science Products). The results indicate that the RFP sequence is present on the ACes.

5 **H. Southern analysis**

Genomic DNA was harvested from the cell lines containing an ACes with the correct recombinant event and digested with *EcoR* I. The digested DNAs were separated on a 0.7% agarose gel, transferred and fixed to a nylon membrane and probed with RFP coding sequences. The
10 result showed that there is an integrated copy of RFP coding sequence in each clone.

EXAMPLE 5

Delivery of a second gene encoding GFP onto the RFP platform ACes

15 **A. Construction of integration vector containing attB and GFP (pD2eGFPIresPuroattB).**

The plasmid pIRESpuo2 (Clontech, Palo Alto, CA; SEQ ID NO: 88) was digested with *EcoRI* and *NotI* then ligated to the D2eGFP *EcoRI-NotI* fragment from pD2eGFP-N1 (Clontech, Palo Alto, CA) to create pD2eGFPIresPuro2. Subsequently, oligos encoding the attB site were
20 annealed and ligated into the *NruI* site of pD2eGFPIresPuro2 to create pD2eGFPIresPuroattB. The orientation of attB in the *NruI* site was determined by PCR.

B. Transfection of LMtk- cells

The LMtk- cells containing the RFP platform ACes produced in
25 Example 4, which has multiple attP sites, were co-transfected with pCXLamIntR and pD2eGFPIresPuroattB using LipofectAMINE PLUS reagent. Five μ g of each vector was placed into a tube containing 750 μ l of DMEM (Dulbecco's modified Eagles Medium). Twenty μ l of the Plus reagent was added to the DNA and incubated at room temperature for 15

-113-

minutes. A mixture of 30 μ l of lipofectamine and 750 μ l DMEM was added to the DNA mixture and incubated an additional 15 minutes at room temperature. The DNA mixture was then added dropwise to approximately 3 million cells attached to a 10cm dish in 5 mls of DMEM.

- 5 The cells were incubated 4 hours (37°C, 5% CO₂) with the DNA-lipid mixture, after which DMEM with 20% fetal bovine serum was added to the dishes to bring the culture medium to 10% fetal bovine serum. The dishes were incubated at 37°C with 5% CO₂.

Plasmid pD2eGFPIresPuroattB has a puromycin gene

- 10 transcriptionally linked to the GFP gene *via* an IRES element. Two days after the transfection the cells were placed in medium containing puromycin at 4 μ g/ml to select for cells containing the pD2eGFPIresPuroattB plasmid integrated into the genome. Twenty-three clones were isolated after 17 days of selection with puromycin. These
15 clones were expanded and then analyzed for the presence of the GFP gene on the ACes by 2-color (RFP/biotin & GFP/digoxigenin) TSA-FISH (NEN) according to the manufacturers protocol. Sixteen of the 23 clones produced a positive FISH signal on the ACes with a GFP probe.

EXAMPLE 6

- 20 **Delivery Of ACes Into human Mesenchymal Stem Cells (hMSC)**

A. Transfection

- Transfection conditions for the most efficient delivery of the ACes into hMSCs (Cambrex BioWhittaker Product Code PT-2501, lot# F0658, East Rutherford, New Jersey) were assayed using LipofectAMINE PLUS
25 and Superfect. One million prototype B ACes, which is a murine derived 60Mb ACes having primarily murine pericentric heterochromatin, and carrying a "payload" containing a hygromycin B selectable marker gene and a *lacZ* reporter gene (see , Telenius et al., 1999, Chrom. Res., 7:3-7; and Kereso et al., 1996, Chrom. Res., 4:226-239; each of which is

-114-

incorporated herein by reference in its entirety), were combined with 1-12 μ l of the transfection agent. In the case of LipofectAMINE PLUS, the PLUS reagent was combined with the ACes for 15 minutes followed by LipofectAMINE for a further 15 minutes. Superfect was complexed for

5 10 minutes at a ratio of 2 μ l Superfect per 1 million ACes. The ACes/transfection agent complex was then applied to 0.5 million recipient cells and the transfection was allowed to proceed according to the manufacturer's protocol. Percent transfected cells was determined on a FACS Vantage flow cytometer with argon laser tuned to 488 nm at

10 200mW and FITC fluorescence collected through a standard FITC 530/30 nm band pass filter. After 24 hours, IdUrd labeled ACes were delivered to human MSCs in the range of 30-50%, varying with transfection agent and dose. ACes delivery curves were generated from data collected in experiments that varied the dose of the transfection reagents. Dose

15 response curves of Superfect and LipofectAMINE PLUS, showing delivery of ACes into recipient hMSCs cells, were prepared, measured by transfer of IdUrd labeled ACes and detected by flow cytometry. Superfect shows maximum delivery in the range of 30-50% at doses greater than 2 μ l per million ACes. LipofectAMINE PLUS has a 42-48% delivery peak around

20 5-8 μ l per million ACes. These dose curves were then correlated with toxicity data to determine the transfection conditions that will allow for highest potential transfection efficiency. Toxicity was determined by a modified plating efficiency assay (de Jong et al., 2001, Chrom. Research, 9:475-485). The population's normalized plating efficiency (at maximum

25 % delivery doses) was in the range of 0.2 - 0.4 for Superfect and 0.5 - 0.6 with LipofectAMINE PLUS.

Due to the transfected population consisting of mixed cell types, flow cytometry allowed for the assessment of ACes delivery into each sub-population and the purification of the target population. Flow profiles

-115-

showing forward scatter (cell size) and side scatter (internal cell granularity) revealed three distinct hMSC populations that were gated into three regions: R3 (small cell region), R4 (medium cell region), R5 (large cell region). Transfection conditions were further optimized by re-analyzing delivery curves and assessing the differences in delivery to each sub-population. Dose response curves of Superfect and LipofectAMINE were prepared showing % delivery to each sub-population represented by the gating on basis of cell size and granularity properties of the mixed population. Three distinct hMSC populations were gated and % delivery dose curves generated. Using Superfect and LipofectAMINE PLUS the overall % delivery increased with cell size (80-90% delivery in large cells). LipofectAMINE PLUS at high doses (8-12 μ l per 1 million ACes) shows an increase in the overall proportion of chromosome transfer to the small population (10-20%). This suggests an advantage to using this transfection agent if the small-undifferentiated cell population is the desired target host cell.

B. Expression from Genes on ACes IN hMSCs

Following the delivery screening process conducted in section (A) above, the most promising results were subjected to further analyses to monitor expression and verify the presence of structurally intact ACes. The transfection conditions employed for these experiments were exactly the same as those that had been used during the screening process. Short-term expression was monitored by transfecting hMSCs with ACes containing a RFP gene (red fluorescent protein) set forth in Example 2C as "D11C4". The unselected population was harvested at 72-96 hours post transfection and % positive fluorescent cells measured by flow cytometry. RFP expression was in the range of 1-20%.

Long term-gene expression was assayed by selecting for hygromycin B resistant cells over a period of 7-10 days. Cytogenetic

-116-

analysis was done to detect presence of intact *ACes* by Fluorescent *In Situ* hybridization (FISH), where metaphase chromosomes were hybridized to a mouse major satellite-DNA probe (targeting murine pericentric heterochromatin) and a lambda probe (hybridizing to the *lacZ* gene). The human mesenchymal transfected culture could not undergo standard sub-cloning as diffuse colonies form with limited doublings available for expansion. Cytogenetic analysis was performed on the entire population, sampling over a period of 3-10 days post-transfection. The hygromycin resistant population was then blocked in mitosis with colchicine and analyzed for presence of intact *ACes* by FISH. Preliminary FISH results show approximately 2-8% of the hMSC-transfected population had an intact *ACes*. This compared to rat skeletal muscle myoblast clones, which were in the range of 60-95%. To increase the % of intact *ACes* in the hMSC-transfected population an enrichment step can be utilized as described in Example 2C.

C. Differentiation of The hMSCs

In initial experiments where transfected hMSCs cells have been induced to differentiate into adipose or osteocytes, the results indicate that the transfected cells appear to be differentiating at a rate comparable to the untransfected controls and the cultures are lineage specific as tested by microscopic examination, FISH, Oil Red O staining (adipocyte assay), and calcium secretion (osteocyte assay).

Accordingly, these results indicate that the artificial chromosomes (*ACes*) provided herein can be successfully transferred into hMSC target cells. Targeting MSCs (such as hMSCs) permits gene transfer into cells in an undifferentiated state where the cells are easier to expand and purify. The genetically modified cells can then be differentiated *in vitro* or injected into a site *in vivo* where the microenvironment will induce transformation into specific cell lineages.

-117-

EXAMPLE 7**Delivery of a Promoterless Marker Gene to a Platform ACes**

Platform ACes containing pSV40attPsensePURO (Figure 4) were constructed as set forth in Examples 3 and 4.

5 A. Construction of Targeting Vectors.

The base vector p18attBZeo (3166bp; SEQ ID NO: 114) was constructed by ligating the 1067bp *HindIII*-*SspI* fragment containing attBZeo, obtained from pLITattBZeo (SEQ ID NO:91), into pUC18 (SEQ ID NO: 122) digested with *HindIII* and *SspI*.

10 1. p18attBZEO-eGFP (6119bp; SEQ ID NO: 126) was constructed by inserting the 2977bp *SpeI*-*HindIII* fragment from pCXeGFP (SEQ ID NO:71; Okabe, *et al.* (1997) *FEBS Lett* 407:313-319) containing the eGFP gene into p18attBZeo (SEQ ID NO: 114) digested with *HindIII* and *XbaI*.

2. p18attBZEO-5'6XHS4eGFP (Figure 10; 7631bp; SEQ ID NO: 116) was constructed by ligating the 4465bp *HindIII* fragment from pCXeGFPattB(6XHS4)2 (SEQ ID NO: 123) which contains the eGFP gene, under the regulation of the chicken beta actin promoter, 6 copies of the HS4 core element located 5' of the chicken beta actin promoter and the polyadenylation signal into the *HindIII* site of p18attBZeo (SEQ ID NO: 15 114).
20 114).

3. p18attBZEO-3'6XHS4eGFP (Figure 11; 7600bp; SEQ ID NO: 115) was created by removing the 5'6XHS4 element from p18attBZeo-(6XHS4)2eGFP (SEQ ID NO: 110). p18attBZeo-(6XHS4)2eGFP was digested with *EcoRV* and *SpeI*, treated with Klenow and religated to form
25 p18attBZeo3'6XHS4eGFP (SEQ ID NO: 115).

4. p18attBZEO-(6XHS4)2eGFP (Figure 12; 9080bp; SEQ ID NO: 110) was created in two steps. First, the *EcoRI*-*SpeI* fragment from pCXeGFPattB(6XHS4)2 (SEQ ID NO: 123) which contains 6 copies of the HS4 core element was ligated into p18attBZeo (SEQ ID NO: 114)

-118-

digested with *EcoRI* and *XbaI* to create p18attBZeo6XHS4 (4615bp; SEQ ID NO: 117). Next, p18attBZeo6XHS4 was digested with *HindIII* and ligated to the 4465bp *HindIII* fragment from pCXeGFPattB(6XHS4)2 which contains the eGFP gene, under the regulation of the chicken beta actin promoter, 6 copies of the HS4 core element located 5' of the chicken beta actin promoter and the polyadenylation signal.

Table 2

Targeting plasmid	No. zeocin resistant clones	No. clones with expected PCR product size	No. clones with correct sequence at recombination junction
p18attBZEO-eGFP	12	12	NT*
p18attBZEO-5'6XHS4eGFP	11	11	NT
p18attBZEO-3'6XHS4eGFP	11	11	NT
p18attBZEO-(6XHS4)2eGFP	9	9	4/4

*NT = not tested

B. Transfection and Selection with Drug.

The mouse cell line containing the 2nd generation platform ACE, B19-38 (constructed as set forth in Example 3), was plated onto four 10cm dishes at approximately 5 million cells per dish. The cells were incubated overnight in DMEM with 10% fetal calf serum at 37°C and 5% CO₂. The following day the cells were transfected with 5μg of each of the 4 vectors listed in Example 7.A. above and 5μg of pCXLamIntR (SEQ ID NO: 112), for a total of 10μg per 10cm dish. Lipofectamine Plus reagent was used to transfect the cells according to the manufacturers protocol. Two days post-transfection zeocin was added to the medium at 500ug/ml. The cells were maintained in selective medium until colonies formed. The colonies were then ring-cloned (see, e.g., McFarland, 2000, Methods Cell Sci, Mar;22(1):63-66).

C. Analysis of Clones (PCR, SEQUENCING).

-119-

Genomic DNA was isolated from each of the candidate clones with the Wizard kit (Promega) and following the manufacturers protocol. The following primer set was used to analyze the genomic DNA isolated from the zeocin resistant clones: 5PacSV40 –

5 CTGTTAATTAAGTGTGGAATGTGTG TCAGTTAGGGTG (SEQ ID NO:76);
Antisense Zeo - TGAACAGGGTCACGTCGTCC (SEQ ID NO:77). PCR amplification with the above primers and genomic DNA from the site-specific integration of any of the 4 zeocin vectors would result in a 673bp PCR product.

10 As set forth in Table 2, of the 4 zeocin resistant candidate clones thusfar analyzed by PCR, all 4 exhibit the correct sequence for a site-specific integration event.

EXAMPLE 8

Integration of a PCR product by site-specific recombination.

15 In this example a gene is integrated onto the platform ACes by site-specific recombination without cloning said gene into a vector.

A. PCR PRIMER DESIGN.

PCR primers are designed to contain an attB site at the 5' end of one of the primers in the primer set. The remaining primers, which could
20 be one or more than one primer, do not contain an attB site, but are complementary to sequences flanking the gene or genes of interest and any associated regulatory sequences. In first example, 2 primers (one containing an attB site) are used to amplify a selective gene such as puromycin.

25 In a second example as shown in Figure 13, the primer set includes primers 1 & 2 that amplify the GFP gene without amplification of an upstream promoter. Primer 1 contains the attB site at the 5' end of the oligo. Primers 3 & 4 are designed to amplify the IRES-blasticidin DNA sequences from the vector pIRESblasticidin. The 5' end of primer 3

-120-

contains sequences complementary to the 5' end of primer 2 such that annealing can occur between 5' ends of the two primers.

B. PCR REACTION AND SUBSEQUENT LIGATION TO CREATE CIRCULAR MOLECULES FROM THE PCR PRODUCT

5 In the first example set forth above in Section A, the two PCR primers are combined with a puromycin DNA template such as pPUR (Clontech), a heat stable DNA polymerase and appropriate conditions for DNA amplification. The resulting PCR product (attB-Puromycin) is then then purified and self-ligated to form a circular molecule.

10 In the second example set forth above in Section A, amplification of the GFP gene and IRES-blasticidin sequences is accomplished by combining primers 1 & 2 with DNA template pD2eGFP and primers 3 & 4 with template pIRESblasticidin under appropriate conditions to amplify the desired template. After initial amplification of the two products (attB-GFP
15 & IRES-blasticidin) in separate reactions, a second round of amplification using both of the PCR products from the first round of amplification together with primers 1 and 4 amplifies the fusion product attB-GFP-IRES-blasticidin (Figure 13). This technique of using complementary sequences in primer design to create a fusion product is employed in *Saccharomyces*
20 *cerevisiae* for allele replacement (Erdeniz *et al* (1997) *Gen Res* 7:1174-1183). The amplified product is then purified from the PCR reaction mixture by standard methods and ligated to form a circular molecule.

C. INTRODUCTION OF PCR PRODUCT ONTO THE ACes USING A RECOMBINASE

25 The circular PCR product is then be introduced to the platform ACes using the bacteriophage lambda integrase E174R. The introduction can be performed *in vivo* by transfecting the pCXLamIntR (SEQ ID NO: 112) vector encoding the lambda integrase mutant E174R together with the circularized PCR product into a cell line containing the platform ACE.

-121-

D. SELECTION FOR MARKER GENE

The marker gene (in this case either puromycin, blasticidin or GFP) is used to enrich the population for cells containing the proper integration event. A proper integration event in the second example (Figure 14) juxtaposes a promoter residing on the platform ACes 5' to the attB-GFP-IRES-Blasticidin PCR product, allowing for transcription of both GFP and blasticidin. If enrichment is done by drug selection, blasticidin is added to the medium on the transfected cells 24-48 hours post-transfection. Selection is maintained until colonies are formed on the plates. If enrichment is done by cell sorting, cells are sorted 2-4 days post-transfection to enrich for cells expressing the fluorescent marker (GFP in this case).

E. ANALYSIS OF CLONES

Clonal isolates are analyzed by PCR, FISH and sequence analysis to confirm proper integration events.

EXAMPLE 9**Construction of a human platform ACes "ACE 0.1"****A. CONSTRUCTION OF THE TARGETING VECTOR pPACrDNA**

Genome Systems (IncyteGenomics) was supplied with the primers 5'HETS (GGGCCGAAACGATCTCAACCTATT; SEQ ID NO:78), and 3'HETS (CGCAGCGGCCCTCCTACTC; SEQ ID NO:79), which were used to amplify a 538bp PCR product homologous to nt 9680-10218 of the human rDNA sequences (GenBank Accession No. U13369) and used as a probe to screen a human genomic P1AC (P1 Artificial Chromosome) library constructed in the vector pCYPAC2 (Ioannou *et al.* (1994) *Nat. Genet.* 6(1): 84-89). Genome Systems clone #18720 was isolated in this screen and contains three repeats of human rDNA as assessed by restriction analysis. GS clone #18720, was digested with PmeI, a restriction enzyme unique to a single repeat of the human rDNA (45Kbp),

-122-

and then religated to form pPACrDNA (Figure 15). The insert in pPACrDNA was analyzed by restriction digests and sequence analysis of the 5' and 3' termini. The pPACrDNA, rDNA sequences are homologous to Genbank Accession #U13369, containing an insert of about 45 kB
5 comprising a single repeat beginning from the end of one repeat at ~33980 (relative to the Genbank sequence) through the beginning of the next repeat up to approximately 35120 (the repeat offset from that listed in the GenBank file). Thus, the rDNA sequence is just over 1 copy of the repeat extending from 33980 (+/-10bp) to the end of the first repeat
10 (43Kbp) and continuing into the second repeat to bp 35120 (+/-10bp).

B. TRANSFECTION AND ACes FORMATION.

Five hundred thousand MSU1.1 cells (Morgan et al., 1991, Exp. Cell Res., Nov;197(1):125-136; provided by Dr. Justin McCormick at Michigan State University) were plated per 6cm plate (3 plates total) and
15 allowed to grow overnight. The cells were 70-80% confluent the following day. One plate was transfected with 15µg pPACrDNA (linearized with *Pme* I) and 2µg pSV40attPsensePuro (linearized with *Sca* I; see Example 3). The remaining plates were controls and were transfected with either 20µg pBS (Stratagene) or 20µg
20 pSV40attBsensePuro (linearized with *Sca* I). All three plates were transfected using a CaPO₄ protocol.

C. SELECTION OF PUROMYCIN RESISTANT COLONIES

One day post-transfection the cells were "glycerol shocked" by the addition of PBS medium containing 10% glycerol for 30 seconds.
25 Subsequently, the glycerol was removed and replaced with fresh DMEM. Four days post-transfection selective medium was added. Selective medium contains 1ug/ml puromycin. The transfection plates were maintained at 37°C with 5% CO₂ in selective medium for 2 weeks at which point colonies could be seen on the plate transfected with

-123-

pPACrDNA and pSV40attPsensePuro. The colonies were ring-cloned from the plate on day 17 post-selection and expanded in selective medium for analysis. Only two colonies (M2-2d & M2-2b) were able to proliferate in the selective medium after cloning. No colonies were seen
5 on the control plates after 37 days in selective medium.

D. ANALYSIS OF CLONES

FISH analysis was performed on the candidate clones to detect ACes formation. Metaphase spreads from the candidate clones were probed in multiple probe combinations. In one experiment, the probes
10 used were biotin-labeled human alphoid DNA (pPACrDNA) and digoxigenin-labeled mouse major DNA (pFK161) as a negative control. Candidate M2-2d was single cell subcloned by flow sorting and the candidate subclones were reanalyzed by FISH. Subclone 1B1 of M2-2d was determined to be a platform ACes and is also designated human
15 Platform ACE 0.1.

EXAMPLE 10

Site-specific integration of a marker gene onto a human platform ACE 0.1

The promoterless delivery method was used to deliver a promoterless blasticidin marker gene onto the human platform ACes with
20 excellent results. The human ACes platform with a promoterless blasticidin marker gene resulted in 21 of 38 blasticidin resistant clones displaying a PCR product of the expected size from the population co-transfected with pLIT38attBSRpolyA10 and pCXLamIntR (Figure 8; SEQ ID NOs. 111 and 112). Whereas, the population transfected with
25 pBlueScript resulted in 0 blasticidin resistant colonies.

A. CONSTRUCTION OF pLIT38attB-BSRpolyA10 & pLIT38attB-BSRpolyA2.

The vector pLITMUS 38 (New England Biolabs; U.S. Patent No. 5,691,140; SEQ ID NO: 119) was digested with *EcoRV* and ligated to

-124-

- two annealed oligomers, which form an attB site (attB1 5'-TGAAGCCTGCTTTTTTATACTAACTTGAGCGAA-3' (SEQ ID NO:8); attB2 5'- TTCGCTCAAGTTAGTATAAAAAAGCAGGCTTCA-3'; SEQ ID NO:9). This ligation reaction resulted in the vector pLIT38attB (SEQ ID NO: 120).
- 5 The blasticidin resistance gene and SV40 polyA site was PCR amplified with primers: 5BSD (ACCATGAAAACATTTAACATTTCTCAACA; SEQ ID NO:80) and SV40polyA (TTTATTTGTGAAATTTGTGATGCTATTGC; SEQ ID NO:81) using pPAC4 (Frengen, E., *et al.* (2000) Genomics 68 (2), 118-126; GenBank Accession No. U75992) as template. The blasticidin-
- 10 SV40polyA PCR product was then ligated into pLIT38attB at the *Bam*HI site, which was Klenow treated following digestion with *Bam*HI. pLIT38attB-BSDpolyA10 (SEQ ID NO: 111) and pLIT38attB-BSDpolyA2 (SEQ ID NO: 121) are the two resulting orientations of the PCR product ligated into the vector.
- 15 **B. TRANSFECTION OF MSU1.1 CELLS CONTAINING HUMAN PLATFORM ACE 0.1.**
- MSU1.1 cells containing human platform ACE 0.1 (see Example 9) was expanded and plated to five 10cm dishes with 1.3×10^6 cells per dish. The cells were incubated overnight in DMEM with 10% fetal bovine
- 20 serum, at 37°C and 5% CO₂. The following day the cells were transfected with 5µg of each plasmid as set forth in Table 3, for a total of 10µg of DNA per plate of cells transfected (see Table 3) using ExGen 500 *in vitro* transfection reagent (MBI fermentas, cat. no. R0511). The transfection was performed according to the manufacturers protocol.
- 25 Cells were incubated at 37°C with 5% CO₂ in DMEM with 10% fetal bovine serum following the transfection.

-125-

Table 3

Plate #	Plasmid 1	Plasmid 2	No. Bsd ^R Colonies
1	pBS	None	0
2	pCXLamInt	pLIT38attB-BSRpolyA10	16
3	pCXLamIntR	pLIT38attB-BSRpolyA10	40
4	pCXLamInt	pLIT38attB-BSRpolyA2	28
5	pCXLamIntR	pLIT38attB-BSRpolyA2	36

10 C. SELECTION OF BLASTICIDIN RESISTANT CLONES.

Three days following the transfection the cells were split from a 10 cm dish to two 15cm dishes. The cells were maintained in DMEM with 10% fetal bovine serum for 4 days in the 15 cm dishes. Seven days post-transfection blasticidin was introduced into the medium. Stably transfected cells were selected with 1 μ g/ml blasticidin. The number of colonies formed on each plate is listed in Table 3. These colonies were ring-cloned and expanded for PCR analysis. Upon expansion in blasticidin containing medium some clones failed to live and therefore do not have corresponding PCR data.

20 D. PCR ANALYSIS

Thirty-eight of the 40 clones from plate 3 grew after ring-cloning. Genomic DNA was isolated from these clones with the Promega Wizard Genomic cDNA purification kit, digested with *Eco*RI and used as template in a PCR reaction with the following primers: 3BSP – TTAATTTCTGGG TATATTTGAGTGGA (SEQ ID NO:82); 5PacSV40 – CTGTTAATTAAGTGTGGAA TGTGTGTCAGTTAGGGTG (SEQ ID NO:76). The PCR conditions were as follows. 100ng of genomic DNA was

-126-

amplified with 0.5ul Herculanase polymerase (Stratagene) in a 50ul reaction that contained 12.5pmole of each primer, 2.5mM of each dNTP, and 1X Herculanase buffer (Stratagene). The reactions were placed in a PerkinElmer thermocycler programmed as follows: Initial denaturation at 95°C for 10 minutes; 35 cycles of 94°C for 1 minute, 53°C for 1 minute, 72°C for 1 minute, and 72°C for 1 minute; Final extension for 10 minutes at 72°C; and 4°C hold. If pLIT38attB-BSRpolyA10 integrates onto the human platform ACE 0.1 correctly, PCR amplification with the above primers should yield an 804bp product. Twenty-one of the 38 clones from plate 3 produced a PCR product of the expected 804bp size.

EXAMPLE 11

Delivery of a Vector comprising a Promoterless Marker Gene and a gene encoding a therapeutic product to a Platform ACes

Platform ACes containing pSV40attPsensePURO (Figure 4) were constructed as set forth in Examples 3 and 4.

A. CONSTRUCTION OF DELIVERY VECTORS

1. Erythropoietin cDNA vector, p18EPOcDNA.

The erythropoietin cDNA was PCR amplified from a human cDNA library (E. Perkins *et al.*, 1999, *Proc. Natl. Acad. Sci. USA* 96(5): 2204-2209) using the following primers: EPO5XBA - TATCTAGAATGGGGGTGC ACGAATGTCCTGCC (SEQ ID NO: 83); EPO3BSI - TACGTACGTCATC TGTCCCCTGTCCTGCAGGC (SEQ ID NO: 84). The cDNA was amplified through two successive rounds of PCR using the following conditions: heat denaturation at 95°C for 3 minutes; 35 cycles of a 30 second denaturation (95°C), 30 seconds of annealing (60°C), and 1 minute extension (72°C); the last cycle is followed by a 7 minute extension at 72°C. BIO-X-ACT (BIOLINE) was used to amplify the erythropoietin cDNA from 2.5ng of the human cDNA library in the first round of amplification. Five µl of the first amplification product was used

-127-

as template for the second round of amplification. Two PCR products were produced from the second amplification with Taq polymerase (Eppendorf), each product was cloned into pCR2.1-Topo (Invitrogen) and sequenced. The larger PCR product contained the expected cDNA
5 sequence for erythropoietin. The erythropoietin cDNA was moved from pTopoEPO into p18attBZeo(6XHS4)2eGFP (SEQ ID NO: 110). pTopoEPO was digested with BsiWI and XbaI to release a 588 bp EPO cDNA. BsrGI and BsiWI create compatible ends. The eGFP gene was removed from p18attBZeo(6XHS4)2eGFP by digestion with BsiWI and XbaI, the 8.3 Kbp
10 vector backbone was gel purified and ligated to the 588 bp EPO cDNA to create p18EPOcDNA (SEQ ID NO: 124).

2. Genomic erythropoietin vector, p18genEPO.

The erythropoietin genomic clone was PCR amplified from a human genomic library (Clontech) using the following primers: GENEPO3BSI -
15 CGTACGTCATCTGTCCCCT GTCCTGCA (SEQ ID NO: 85); GENEPO 5XBA -TCTAGAATGGGGGT GCACGGTGAGTACT (SEQ ID NO: 86). The reaction conditions for the amplification were as follows: heat denaturation for 3 minutes (95°C); 30 cycles of a 30 second denaturation (95°C), 30 seconds annealing (from 65°C decreasing 0.5°C per cycle to
20 50°C), and 3 minutes extension (72°C); 15 cycles of a 30 second denaturation (95°C), 30 seconds annealing (50°C), and 3 minute extension (72°C); the last cycle is followed by a 7 minute extension at 72°C. The erythropoietin genomic PCR product (2147 bp) was gel purified and cloned into pCR2.1Topo to create pTopogenEPO. Sequence
25 analysis revealed 2bp substitutions and insertions in the intronic sequences of the genomic clone of erythropoietin. A partial digest with XbaI and complete digest with BsiWI excised the erythropoietin genomic insert from pTopogenEPO. The resulting 2158 bp genomic erythropoietin fragment was ligated into the 8.3 Kbp fragment resulting from the

-128-

digestion of p18attBZeo(6XHS4)2eGFP (SEQ ID NO: 110) with XbaI and BsrGI to create p18genEPO (SEQ ID NO: 125).

B. TRANSFECTION AND SELECTION WITH DRUG

The erythropoietin genomic and cDNA genes were each moved
5 onto the platform *ACes* B19-38 (constructed as set forth in Example 3) by
co-transfecting with pCXLamIntR. Control transfections were also
performed using pCXLamInt (SEQ ID NO: 127) together with either
p18EPOcDNA (SEQ ID NO: 124) or p18genEPO (SEQ ID NO: 125).
Lipofectamine Plus was used to transfect the DNA's into B19-38 cells
10 according to the manufacturer's protocol. The cells were placed in
selective medium (DMEM with 10% FBS and Zeocin @ 500ug/ml) 48
hours post-transfection and maintained in selective medium for 13 days.
Clones were isolated 15 days post-transfection.

C. ANALYSIS OF CLONES (ELISA, PCR)

15 1. ELISA Assays

Thirty clones were tested for erythropoietin production by an ELISA
assay using a monoclonal anti-human erythropoietin antibody (R&D
Systems, Catalogue # MAB287), a polyclonal anti-human erythropoietin
antibody (R & D Systems, Catalogue # AB-286-NA) and alkaline
20 phosphatase conjugated goat-anti-rabbit IgG (heavy and light chains)
(Jackson ImmunoResearch Laboratories, Inc., Catalogue # 111-055-144).
The negative control was a Zeocin resistant clone isolated from B19-38
cells transfected with p18attBZeo(6XHS4) (SEQ ID NO: 117; no insert
control vector) and pCXLamIntR (SEQ ID NO: 112). The preliminary
25 ELISA assay was executed as follows: 1) Nunc-Immuno Plates (MaxiSorb
96-well, Catalogue # 439454) were coated with 75ul of a 1/200 dilution
(in Phosphate buffered Saline, pH 7.4 (PBS), Sigma Catalogue # P-3813)
of monoclonal anti-human erythropoietin antibody overnight at 4°C. 2)
The following day the plates were washed 3 times with 300ul PBS

-129-

containing 0.15% Tween 20 (Sigma, Catalogue # P-9416). 3) The plates were then blocked with 300ul of 1% Bovine Serum Albumin (BSA; Sigma Catalogue # A-7030) in PBS for 1 hour at 37°C. 4) Repeat the washes as in step 2. 5) The clonal supernatants (75ul per clone per well of 96-well plate) were then added to the plate and incubated for 1 hour at 37°C. The clonal supernatant analyzed in the ELISA assay had been maintained on the cells 7 days prior to analysis. 6) Repeat the washes of step 2. 7) Add 75ul of polyclonal anti-human erythropoietin antibody (1/250 dilution in dilution buffer (0.5% BSA, 0.01% Tween 20, 1X PBS, pH 7.4) and incubate 1 hour at 37°C. 8) Repeat washes of step 2. 9) Add 75ul of goat anti-rabbit conjugated alkaline phosphatase diluted 1/4000 in dilution buffer and incubate 1 hour at 37°C. 10) Repeat washes of step 2. 11) Add 75ul substrate, p-nitrophenyl phosphate (Sigma N2640), diluted to 1mg/ml in substrate buffer (0.1 Ethanolamine-HCl (Sigma, Catalogue # E-6133), 5mM MgCl₂ (Sigma, Catalogue # M-2393), pH 9.8). Incubate the plates in the dark for 1 hour at room temperature (22°C). 12) Read the absorption at 405nm (reference wavelength 495nm) on an Universal Microplate Reader (Bio-Tek Instruments, Inc., model # ELX800 UV). The erythropoietin standard curve was derived from readings of diluted human recombinant Erythropoietin (Roche, catalogue # 1-120-166; dilution range 125 - 7.8mUnits/ml). From this preliminary assay the 21 clones displaying the highest expression of erythropoietin were analyzed a second time in the same manner using medium supernatants that had been on the clones for 24 hours and a 1:3 dilution thereof.

2. PCR Analysis

Genomic DNA was isolated from the 21 clones with the best expression (as assessed by the initial ELISA assay above) as well as the B19-38 cell line and used for PCR analysis. Genomic DNA was isolated using the Wizard genomic DNA purification kit (Promega) according to the

-130-

manufacturers protocol. Amplification was performed on 100ng of genomic DNA as template with MasterTaq DNA Polymerase (Eppendorf) and the primer set 5PacSV40 – CTGTTAATTAAGTGTGGAATGTGTG TCAGTTAGGGTG (SEQ ID NO: 76) and Antisense Zeo -

- 5 TGAACAGGGTCACGTCGTCC (SEQ ID NO: 77). The amplification conditions were as follows: heat denaturation for 3 minutes (95°C); 30 cycles of a 30 second denaturation (95°C), 30 seconds annealing (from 65°C decreasing 0.5°C per cycle to 50°C), and 1 minutes extension (72°C); 15 cycles of a 30 second denaturation (95°C), 30 seconds
- 10 annealing (50°C), and 1 minute extension (72°C); the last cycle is followed by a 10 minute extension at 72°C. PCR products were size separated by gel electrophoresis. Of the 21 clones analyzed 19 produced a PCR product of 650 bp as expected for a site-specific integration event. All nineteen clones were the result of transformations with p19EPOcDNA
- 15 (5) or p18genEPO (14) and pCXLamIntR (i.e. mutant integrase). The remaining two clones, both of which were the result of transformation with p18genEPO (SEQ ID NO: 125) and pCXLamInt (i.e. wildtype integrase; SEQ ID NO: 127), produced a 400 bp PCR product.

Example 12

20 Preparation of a Transformation Vector Useful for the Induction of Plant Artificial Chromosome Formation

- Plant artificial chromosomes (PACs) can be generated by introducing nucleic acid, such as DNA, which can include a targeting DNA, for example rDNA or lambda DNA, into a plant cell, allowing the cell
- 25 to grow, and then identifying from among the resulting cells those that include a chromosome with a structure that is distinct from that of any chromosome that existed in the cell prior to introduction of the nucleic acid. The structure of a PAC reflects amplification of chromosomal DNA, for example, segmented, repeat region-containing and heterochromatic

-131-

structures. It is also possible to select cells that contain structures that are precursors to PACs, for example, chromosomes containing more than one centromere and/or fragments thereof, and culture and/or manipulate them to ultimately generate a PAC within the cell.

5 In the method of generating PACs, the nucleic acid can be introduced into a variety of plant cells. The nucleic acid can include targeting DNA and/or a plant expressible DNA encoding one or multiple selectable markers (*e.g.*, DNA encoding bialophos (bar) resistance) or scorable markers (*e.g.*, DNA encoding GFP). Examples of targeting DNA
10 include, but are not limited to, *N. tabacum* rDNA intergenic spacer sequence (IGS) and *Arabidopsis* rDNA such as the 18S, 5.8S, 26S rDNA and/or the intergenic spacer sequence. The DNA can be introduced using a variety of methods, including, but not limited to *Agrobacterium*-mediated methods, PEG-mediated DNA uptake and electroporation using,
15 for example, standard procedures according to Hartmann *et al* [(1998) *Plant Molecular Biology* 36:741]. The cell into which such DNA is introduced can be grown under selective conditions and can initially be grown under non-selective conditions and then transferred to selective media. The cells or protoplasts can be placed on plates containing a
20 selection agent to grow, for example, individual calli. Resistant calli can be scored for scorable marker expression. Metaphase spreads of resistance cultures can be prepared, and the metaphase chromosomes examined by FISH analysis using specific probes in order to detect amplification of regions of the chromosomes. Cells that have artificial
25 chromosomes with functioning centromeres or artificial chromosomal intermediate structures, including, but not limited to, dicentric chromosomes, formerly dicentric chromosomes, minichromosomes, heterochromatin structures (*e.g.* sausage chromosomes), and stable self-replicating artificial chromosomal intermediates as described herein, are

-132-

identified and cultured. In particular, the cells containing self-replicating artificial chromosomes are identified.

The DNA introduced into a plant cell for the generation of PACs can be in any form, including in the form of a vector. An exemplary
5 vector for use in methods of generating PACs can be prepared as follows.

For the production of artificial chromosomes, plant transformation vectors, as exemplified by pAgIIa and pAgIIb, containing a selectable marker, a targeting sequence, and a scorable marker were constructed using procedures well known in the art to combine the various fragments.
10 The vectors can be prepared using vector pAg1 as a base vector and inserting the following DNA fragments into pAg1: DNA encoding β -glucoronidase under the control of the nopaline synthase (NOS) promoter fragment and flanked at the 3' end by the NOS terminator fragment, a fragment of mouse satellite DNA and an *N. tabacum* rDNA intergenic
15 spacer sequence (IGS). In constructing plant transformation vectors, vector pAg2 can also be used as the base vector.

1. Construction of pAG1

Vector pAg1 (SEQ. ID. NO: 89) is a derivative of the CAMBIA vector named pCambia 3300 (Center for the Application of Molecular
20 Biology to International Agriculture, i.e., CAMBIA, Canberra, Australia; www.cambia.org), which is a modified version of vector pCambia 1300 to which has been added DNA from the bar gene conferring resistance to phosphinothricin. The nucleotide sequence of pCambia 3300 is provided in SEQ. ID. NO: 90. pCambia 3300 also contains a lacZ alpha sequence
25 containing a polylinker region.

pAg1 was constructed by inserting two new functional DNA fragments into the polylinker of pCambia 3300: one sequence containing an attB site and a promoterless zeomycin resistance-encoding DNA flanked at the 3' end by a SV40 polyA signal sequence, and a second

-133-

sequence containing DNA from the hygromycin resistance gene (hygromycin phosphotransferase) conferring resistance to hygromycin for selection in plants. Although the zeomycin-SV40 polyA signal fusion is not expected to function in plant cells, it can be activated in mammalian
5 cells by insertion of a functional promoter element into the attB site by site-specific recombination catalyzed by the Lambda att integrase. Thus, the inclusion of the attB-zeomycin sequences allows for evaluation of functionality of plant artificial chromosomes in mammalian cells by activation of the zeomycin resistance-encoding DNA, and provides an att
10 site for further insertion of new DNA sequences into plant artificial chromosomes formed as a result of using pAg1 for plant transformation. The second functional DNA fragment allows for selection of plant cells with hygromycin. Thus, pAg1 contains DNA from the bar gene conferring resistance to phosphinothricin, DNA from the hygromycin resistance gene,
15 both resistance-encoding DNAs under the control of a separate cauliflower mosaic virus (CaMV) 35S promoter, and the attB-promoterless zeomycin resistance-encoding DNA.

pAg1 is a binary vector containing *Agrobacterium* right and left T-DNA border sequences for use in *Agrobacterium*-mediated transformation
20 of plant cells or protoplasts with the DNA located between the border sequences. pAg1 also contains the pBR322 Ori for replication in *E.coli*. pAg1 was constructed by ligating *HindIII/PstI*-digested p3300attBZeo with *HindIII/PstI*-digested pBSCaMV35SHyg as follows.

a. Generation of p3300attBZeo

25 Plasmid pCambia 3300 was digested with *PstI/EcoRI* and ligated with *PstI/StuI*-digested pLITattBZeo (the nucleotide sequence of pLITattBZeo is provided in SEQ. ID. NO: 91. (containing DNA encoding the zeocin resistance gene and an attB Integrase recognition sequence) to generate p3300attBZeo which contains an attB site, a promoterless

-134-

zeomycin resistance-encoding DNA flanked at the 3' end by a SV40 polyA signal, and a reconstructed *Pst*I site.

b. Generation of pBSCaMV35SHyg

A DNA fragment containing DNA encoding hygromycin

- 5 phosphotransferase flanked by the CaMV 35S promoter and the CaMV 35S polyA signal sequence was obtained by PCR amplification of plasmid pCambia 1302 (GenBank Accession No. AF234298 and SEQ. ID. NO: 92). The primers used in the amplification reaction were as follows:

CaMV35SpolyA:

- 10 5'-CTGAATTAACGCCGAATTAATTCGGGGGATCTG-3' SEQ. ID. NO: 93
CaMV35Spr:

5'-CTAGAGCAGCTTGCCAACATGGTGGAGCA-3' SEQ. ID. NO: 94

The 2100-bp PCR fragment was ligated with *Eco*RV-digested pBluescript II SK+ (Stratagene, La Jolla, CA, U.S.A.) to generate pBSCaMV35SHyg.

15 **c. Generation of pAg1**

- To generate pAg1, pBSCaMV35SHyg was digested with *Hind*III/*Pst*I and ligated with *Hind*III/*Pst*I-digested p3300attBZeo. Thus, pAg1 contains the pCambia 3300 backbone with DNA conferring resistance to phosphinothricin and hygromycin under the control of
- 20 separate CaMV 35S promoters, an attB-promoterless zeomycin resistance-encoding DNA recombination cassette and unique sites for adding additional markers, *e.g.*, DNA encoding GFP. The attB site can be used as described hereing for the addition of new DNA sequences to plant artificial chromosomes, including PACs formed as a result of using the
- 25 pAg1 vector, or derivatives thereof, in the production of PACs. The attB site provides a convenient site for recombinase-mediated insertion of DNAs containing a homologous att site.

2. pAG2

-135-

The vector pAg2 (SEQ. ID. NO: 95) is a derivative of vector pAg1 formed by adding DNA encoding a green fluorescent protein (GFP), under the control of a NOS promoter and flanked at the 3' end by a NOS polyA signal, to pAg1. pAg2 was constructed as follows. A DNA fragment
5 containing the NOS promoter was obtained by digestion of pGEM-T-NOS, or pGEMEasyNOS (SEQ. ID. NO: 96), containing the NOS promoter in the cloning vector pGEM-T-Easy (Promega Biotech, Madison, WI, U.S.A.), with *Xba*I/*Nco*I and was ligated to an *Xba*I/*Nco*I fragment of pCambia 1302 containing DNA encoding GFP (without the CaMV 35S promoter) to
10 generate p1302NOS (SEQ. ID. NO: 97) containing GFP-encoding DNA in operable association with the NOS promoter. Plasmid p1302NOS was digested with *Sma*I/*Bsi*WI to yield a fragment containing the NOS promoter and GFP-encoding DNA. The fragment was ligated with *Pme*I/*Bsi*WI-digested pAg1 to generate pAg2. Thus, pAg2 contains DNA
15 from the bar gene conferring resistance to phosphinothricin, DNA conferring resistance to hygromycin, both resistance-encoding DNAs under the control of a cauliflower mosaic virus 35S promoter, DNA encoding kanamycin resistance, a GFP gene under the control of a NOS promoter and the attB-zeomycin resistance-encoding DNA. One of skill in
20 the art will appreciate that other fragments can be used to generate the pAg1 and pAg2 derivatives and that other heterologous DNA can be incorporated into pAg1 and pAg2 derivatives using methods well known in the art.

3. pAgIIa and pAgIIb transformation vectors

25 Vectors pAgIIa and pAgIIb were constructed by inserting the following DNA fragments into pAg1: DNA encoding β -glucuronidase, the nopaline synthase terminator fragment, the nopaline synthase (NOS) promoter fragment, a fragment of mouse satellite DNA and an *N. tabacum*

-136-

rDNA intergenic spacer sequence (IGS). The construction of pAgIIa and pAgIIb was as follows.

An *N. tabacum* rDNA intergenic spacer (IGS) sequence (SEQ. ID. NO: 98; see also GenBank Accession No. Y08422; see also Borysyuk *et al.* (2000) *Nature Biotechnology* 18:1303-1306; Borysyuk *et al.* (1997) *Plant Mol. Biol.* 35:655-660; U.S. Patent Nos. 6,100,092 and 6,355,860) was obtained by PCR amplification of tobacco genomic DNA. The IGS can be used as a targeting sequence by virtue of its homology to tobacco rDNA genes; the sequence is also an amplification promoter sequence in plants. This fragment was amplified using standard PCR conditions (*e.g.*, as described by Promega Biotech, Madison, WI, U.S.A.) from tobacco genomic DNA using the primers shown below:

NTIGS-FI

5'- GTG CTA GCC AAT GTT TAA CAA GAT G- 3' (SEQ ID No. 99) and

15 NTIGS-RI

5'-ATG TCT TAA AAA AAA AAA CCC AAG TGA C- 3' (SEQ ID No. 100)

Following amplification, the fragment was cloned into pGEM-T Easy to give pIGS-I. A fragment of mouse satellite DNA (Msat1 fragment; GenBank Accession No. V00846; and SEQ ID No. 101) was amplified via

20 PCR from pSAT-1 using the following primers:

MSAT-F1

5'- AAT ACC GCG GAA GCT TGA CCT GGA ATA TCG C -3'(SEQ ID No. 102) and

MSAT-Ri

25 5'-ATA ACC GCG GAG TCC TTC AGT GTG CA T- 3' (SEQ ID No. 103)

This amplification added a *SacII* and a *HindIII* site at the 5' end and a *SacII* site at the 3' end of the PCR fragment. This fragment was then cloned into the *SacII* site in pIGS-1 to give pMIGS-1, providing a eukaryotic

-137-

centromere-specific DNA and a convenient DNA sequence for detection via FISH.

A functional marker gene containing a NOS-promoter:GUS:NOS terminator fusion was then constructed containing the NOS promoter
5 (GenBank Accession No. U09365; SEQ ID No. 104), *E. coli* β -glucuronidase coding sequence (from the GUS gene; GenBank Accession No. S69414; and SEQ ID No. 105), and the nopaline synthase terminator sequence (GenBank Accession No. U09365; SEQ ID No. 107). The NOS promoter in pGEM-T-NOS was added to a promoterless GUS
10 gene in pBlueScript (Stratagene, La Jolla, CA, U.S.A.) using *NotI/Spel* to form pNGN-1, which has the NOS promoter in the opposite orientation relative to the GUS gene.

pMIGS-1 was digested with *NotI/Spel* to yield a fragment containing the mouse major satellite DNA and the tobacco IGS which was
15 then added to *NotI*-digested pNGN-1 to yield pNGN-2. The NOS promoter was then re-oriented to provide a functional GUS gene, yielding pNGN-3, by digestion and religation with *Spel*. Plasmid pNGN-3 was then digested with *HindIII*, and the *HindIII* fragment containing the β -glucuronidase coding sequence and the rDNA intergenic spacer, along with the Msat
20 sequence, was added to pAG-1 to form pAgIIa (SEQ ID NO: 108), using the unique *HindIII* site in pAg1 located near the right T-DNA border of pAg1, within the T-DNA region.

Another plasmid vector, referred to as pAgIIb, was also recovered, which contained the inserted *HindIII* fragment (SEQ ID NO: 108) in the
25 opposite orientation relative to that observed in pAgIIa. Thus, pAgIIa and pAgIIb differ only in the orientation of the *HindIII* fragment containing the mouse major satellite sequence, the GUS DNA sequence and the IGS sequence. The nucleotide sequences of pAgIIa is provided in SEQ. ID. NOS: 109.

-138-

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

-139-

WHAT IS CLAIMED IS:

1. A eukaryotic chromosome comprising one or a plurality of *att* site(s), wherein:
 - an *att* site is heterologous to the chromosome; and
 - 5 an *att* site permits site-directed integration in the presence of lambda integrase.
2. The eukaryotic chromosome of claim 1, wherein the *att* sites are selected from the group consisting of *attP* and *attB* or *attL* and *attR*, or variants thereof.
- 10 3. The eukaryotic chromosome of claim 1 that is an artificial chromosome.
4. The eukaryotic chromosome of claim 1 that is an artificial chromosome expression system (*ACes*).
5. The eukaryotic chromosome of claim 4 that is predominantly
15 heterochromatin.
6. The chromosome of claim 1 that is an artificial chromosome that contains no more than about 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.
7. The chromosome of claim 1 that is a plant chromosome.
- 20 8. The chromosome of claim 1 that is an animal chromosome.
9. The chromosome of claim 7 that is a plant artificial chromosome.
10. The chromosome of claim 8 that is an animal artificial chromosome.
- 25 11. The chromosome of claim 8 that is a mammalian chromosome.
12. The chromosome of claim 11 that is a mammalian artificial chromosome.

-140-

13. The chromosome of claim 6 that is an artificial chromosome expression system (*ACes*).
14. A platform artificial chromosome expression system (*ACes*) comprising one or a plurality of sites that participate in recombinase
5 catalyzed recombination.
15. The *ACes* of claim 14 that contains one site.
16. The *ACes* of claim 14 that is predominantly heterochromatin.
17. The *ACes* of claim 14 that contains no more than about 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.
- 10 18. The *ACes* of claim 14 that is a plant *ACes*.
19. The *ACes* of claim 14 that is an animal *ACes*.
20. The *ACes* of claim 14 that is selected from a fish, insect, reptile, amphibian, arachnid or a mammalian *ACes*.
21. The *ACes* of claim 14 that is a fish *ACes*.
- 15 22. The artificial chromosome expression system (*ACes*) of claim 14, wherein the recombinase and site(s) are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase system of phage Mu, the Cin recombinase system, the Pin recombinase system of *E. coli* and the R/RS
20 system of the pSR1 plasmid, or any combination thereof.
23. A method of introducing heterologous nucleic acid into a chromosome, comprising:
- contacting a chromosome of any of claims 1 or 14 with a nucleic acid molecule comprising both the heterologous nucleic acid and a
25 recombination site, in the presence of a recombinase that promotes recombination between the sites in the chromosome and in the nucleic acid molecule.

-141-

24. The method of claim 23, wherein the recombinase is selected from the group consisting of Cre, Gin, Cin, Pin, FLP, a phage integrase and R from the pSR1 plasmid.

25. The method of claim 23, wherein the nucleic acid molecule
5 encodes a therapeutic protein, antisense nucleic acid, or comprises an artificial chromosome.

26. The method of claim 25, wherein the nucleic acid molecule comprises a yeast artificial chromosomes (YAC), a bacterial artificial chromosome (BAC) or an insect artificial chromosome (IAC).

10 27. A combination, comprising, the chromosome of claim 1 and a first vector comprising the cognate recombination site, wherein the cognate recombination site is a site that recombines with the site engineered into the chromosome.

28. The combination of claim 27, further comprising nucleic acid
15 encoding a recombinase, wherein the nucleic acid is on a second vector or on the first vector, or on the ACes under an inducible promoter.

29. The combination of claim 28, wherein the recombinase and sites are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase
20 system of phage Mu, the Pin recombinase system of *E. coli* and the R/RS system of the pSR1 plasmid, or any combination thereof.

30. The combination of claim 28, wherein a vector is the plasmid pCXLamIntR.

31. The combination of claim 27, wherein a vector is the plasmid
25 pDsRedN1-attB.

32. A kit, comprising the combination of claim 27 and optionally instructions for introducing heterologous nucleic acid into the chromosome.

-142-

33. A method for introducing heterologous nucleic acid into a platform artificial chromosome, comprising:

(a) mixing an artificial chromosome comprising at least a first recombination site and a vector comprising at least a second

5 recombination site and the heterologous nucleic acid;

(b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the artificial chromosome.

10 34. The method of claim 33, wherein the artificial chromosome is an ACes.

35. The method of claim 33, wherein said mixing step (a) is conducted in cells ex vivo.

15 36. The method of claim 33, wherein said mixing step (a) is conducted extracellularly in an in vitro reaction mixture.

37. The method of claim 33, wherein the at least one recombination protein is encoded by a bacteriophage selected from the group consisting of bacteriophage lambda, phi 80, P22, P2, 186, P4 and P1.

20 38. The method of claim 37, wherein the at least one recombination protein is encoded by bacteriophage lambda, or mutants thereof.

39. The method of claim 33, wherein at least one recombination protein is selected from the group consisting of Int, IHF, Xis and Cre, $\gamma\delta$,
25 Tn3 resolvase, Hin, Gin, Cin and Flp.

40. The method of claim 32, wherein the recombination sites are selected from the group consisting of att and lox P sites.

-143-

41. The method of claim 33, wherein the first and/or second recombination site contains at least one mutation that removes one or more stop codons.

42. The method of claim 33, wherein the first and/or second recombination site contains at least one mutation that avoids hairpin formation.

43. The method of claim 33, wherein the first and/or second recombination site comprises at least a first nucleic acid sequence selected from the group consisting of SEQ ID NOs:41-56:

- 10 a) RKYCWGCTTTYKTRTACNAASTSGB (m-att) (SEQ ID NO:41);
- b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:42);
- c) GTTCAGCTTTCKTRTACNAACTSGB (m-attR) (SEQ ID NO:43);
- d) AGCCWGCTTTCKTRTACNAAAGTSGB (m-attL) (SEQ ID NO:44);
- e) GTTCAGCTTTYKTRTACNAAAGTSGB (m-attP1) (SEQ ID NO:45);
- 15 f) AGCCTGCTTTTTGTACAAACTTGT (attB1) (SEQ ID NO:46);
- g) AGCCTGCTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:47);
- h) ACCCAGCTTCTTGTACAAACTTGT (attB3) (SEQ ID NO:48);
- i) GTTCAGCTTTTTGTACAAACTTGT (attR1) (SEQ ID NO:49);
- j) GTTCAGCTTCTTGTACAAACTTGT (attR2) (SEQ ID NO:50);
- 20 k) GTTCAGCTTCTTGTACAAAGTTGG (attR3) (SEQ ID NO:51);
- l) AGCCTGCTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:52);
- m) AGCCTGCTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:53);
- n) ACCCAGCTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:54);
- o) GTTCAGCTTTTTGTACAAAGTTGG (attP1) (SEQ ID NO:55);
- 25 p) GTTCAGCTTCTTGTACAAAGTTGG (attP2, P3) (SEQ ID NO: 56);

and a corresponding or complementary DNA or RNA sequence, wherein R=A or G, K=G or T/U, Y=C or T/U, W=A or T/U, N=A or C or G or T/U, S=C or G, and B=C or G or T/U; and

-144-

the core region does not contain a stop codon in one or more reading frames.

44. The method of claim 33, wherein the first and/or second recombination site comprises at least a first nucleic acid sequence
5 selected from the group consisting of a mutated att recombination site containing at least one mutation that enhances recombinational specificity, a complementary DNA sequence thereto, and an RNA sequence corresponding thereto.

45. The method of claim 33, wherein the vector comprising the
10 second site further encodes at least one selectable marker.

46. The method of claim 45, wherein the marker is a promoterless marker, which, upon recombination is under the control of a promoter and is thereby expressed.

47. The method of claim 46, wherein the first recombination site
15 is attP and is in the sense orientation prior to recombination.

48. The method of claim 46, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic, fluorescent, or capable of being bound by an antibody and FACs sorted.

20 49. The method of claim 48, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase (hisD).

50. A cell comprising, the chromosome of claim 1.

25 51. The cell of claim 50, wherein the cell is a nuclear donor cell.

52. The cell of claim 50, wherein the cell is a stem cell.

53. The stem cell of claim 52, wherein said stem cell is human and is selected from the group consisting of a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell and an embryonic stem cell.

-145-

54. The cell of claim 50, wherein the cell is mammalian.
55. The cell of claim 54, wherein the mammal is selected from the group consisting of humans, primates, cattle, pigs, rabbits, goats, sheep, mice, rats, guinea pigs, hamsters, cats, dogs, and horses.
- 5 56. The cell of claim 50, wherein the cell is a plant cell.
57. A cell comprising the platform *ACes* of claim 14.
58. The cell of claim 57, wherein the cell is a nuclear donor cell.
59. The cell of claim 57, wherein the cell is a stem cell.
60. The stem cell of claim 59, wherein said stem cell is human
- 10 and is selected from the group consisting of a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell and an embryonic stem cell.
61. A human mesenchymal cell comprising an artificial chromosome.
62. The human mesenchymal cell of claim 61, wherein said
- 15 artificial chromosome is an *ACes*.
63. The human mesenchymal cell of claim 62, wherein the *ACes* is a platform-*ACes*.
64. A method for introducing heterologous nucleic acid into the mesenchymal cell of claim 63, comprising:
- 20 (a) introducing into the cell of claim 63, wherein the platform-*ACes* has a first recombination site, a vector comprising at least a second recombination site and the heterologous nucleic acid;
- (b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between
- 25 the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the platform-*ACes* within the mesenchymal cell.
65. A lambda-intR mutein comprising a glutamic acid to arginine change at position 174 of wild-type lambda-intR.

-146-

66. The lambda-intR mutin of claim 65, wherein the lambda-intR mutin comprises SEQ ID NO:37.

67. The method of claim 46 wherein the promoterless marker is transcriptionally downstream of the heterologous nucleic acid, wherein
5 the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.

68. The method of claim 67, wherein the selectable marker and the heterologous nucleic acid are transcriptionally linked by the presence
10 of a IRES between them.

69. The method of claim 68, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.

15 70. The method of claim 69, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase.

71. The method of claim 67 further comprising expressing the
20 heterologous protein and isolating the heterologous protein.

72. A method for producing a transgenic animal, comprising introducing a platform-ACes into an embryonic cell.

73. The method of claim 72, wherein the embryonic cell is a stem cell.

25 74. The method of claim 72, wherein the embryonic cell is in an embryo.

75. The method of claim 72, wherein the platform-ACes comprises heterologous nucleic acid that encodes a therapeutic product.

-147-

76. The method of claim 72, wherein the transgenic animal is a fish, insect, reptile, amphibians, arachnid or mammal.

77. The method of claim 72, wherein the *ACes* is introduced by cell fusion, lipid-mediated transfection by a carrier system, microinjection,
5 microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer.

78. A transgenic animal produced by the method of claim 72.

79. A cell line useful for making a library of *ACes*, comprising a multiplicity of heterologous recombination sites randomly integrated
10 throughout the endogenous chromosomes.

80. A method of making a library of *ACes* comprising random portions of a genome, comprising introducing one or more *ACes* into the cell line of claim 79, under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of
15 the heterologous recombination sites within the cell's chromosomal DNA; and isolating said multiplicity of *ACes*, thereby producing a library of *ACes* whereby multiple *ACes* have different portions of the genome within.

81. A library of cells useful for genomic screening, said library
20 comprising a multiplicity of cells, wherein each cell comprises an *ACes* having a mutually exclusive portion of a chromosomal nucleic acid therein.

82. The library of cells of claim 81, wherein the cells of the library are from a different species than the chromosomal nucleic acid
25 within the *ACes*.

83. A method of making one or more cell lines, comprising
a) integrating into endogenous chromosomal DNA of a selected cell species, a multiplicity of heterologous recombination sites,

-148-

b) introducing a multiplicity of *ACes* under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites integrated within the cell's endogenous chromosomal DNA;

5 c) isolating said multiplicity of *ACes*, thereby producing a library of *ACes* whereby a multiplicity of *ACes* have mutually exclusive portions of the endogenous chromosomal DNA therein;

d) introducing the isolated multiplicity of *ACes* of step c) into a multiplicity of cells, thereby creating a library of cells;

10 e) selecting different cells having mutually exclusive *ACes* therein and clonally expanding or differentiating said different cells into clonal cell cultures, thereby creating one or more cell lines.

84. The method of claim 23, wherein the nucleic acid molecule with a recombination site is a PCR product.

15 85. Method of claim 23 wherein the recombinase is a protein and the recombination event occurs in vitro.

86. The method of claim 33, wherein the vector is a PCR product comprising a second recombination site.

20 87. The lambda-intR mutein of claim 65, wherein the mutein further comprises an amino acid signal for nuclear localization.

88. The lambda-intR mutein of claim 65, wherein the mutein further comprises an epitope tag for protein purification.

89. A modified iron-induced promoter comprising SEQ ID NO:128.

25 90. A plasmid or expression cassette comprising the promoter of claim 89.

91. A vector, comprising:
a recognition site for recombination; and

-149-

a sequence of nucleotides that targets the vector to an amplifiable region of a chromosome.

92. The vector of claim 91, wherein the amplifiable region comprises heterochromatic nucleic acid.

5 93. The vector of claim 91, wherein the amplifiable region comprises rDNA.

94. The vector of claim 93, wherein the rDNA comprises an intergenic spacer.

95. The vector of claim 91, further comprising nucleic acid
10 encoding a selectable marker that is not operably associated with any promoter.

96. The vector of claim 91, wherein the chromosome is a mammalian chromosome.

97. The vector of claim 91, wherein the chromosome is a plant
15 chromosome.

98. A cell of claim 57 that is a plant cell, wherein the ACes platform is a MAC.

99. The plant cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

20 100. The plant cell of claim 99, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

101. A cell of claim 57 that is an animal cell, wherein the ACes platform is a plant artificial chromosome (PAC).

25 102. The cell of claim 101 that is a mammalian cell.

103. The cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

104. The cell of claim 102, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

-150-

105. The cell of claim 104, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

106. A method, comprising:

- 5 introducing a vector of claim 91 into a cell;
 growing the cells; and
 selecting a cell comprising an artificial chromosome that comprises one or more repeat regions.

107. The method of claim 106, wherein sufficient portion of the
10 vector integrates into a chromosome in the cell to result in amplification of chromosomal DNA.

108. The method of claim 106, wherein the artificial chromosome is an *ACes*.

109. A method for screening, comprising:

- 15 contacting a cell comprising a reporter *ACes* with test compounds or known compounds, wherein:

 the reporter *ACes* comprises one or a plurality of reporter constructs;

- a reporter construct comprises a reporter gene in operative linkage
20 with a regulatory region responsive to test or known compounds; and

 detecting any increase or decrease in signal output from the reporter, wherein a change in the signal is indicative of activity of the test or known compound on the regulatory region.

110. The method of claim 109, wherein the reporter is operatively
25 linked to a promoter that controls expression of a gene in a signal transduction pathway, whereby activation or reduction in the signal indicates that the pathway is activated or down-regulated by the test compound.

-151-

111. The method of claim 109, wherein the reporter in the construct encodes drug resistance or encodes a fluorescent protein.

112. The method of claim 111, wherein the fluorescent protein is selected from the group consisting of red, green and blue fluorescent
5 proteins.

113. The method of claim 109, wherein the *ACes* comprises a plurality of reporter-linked constructs, each with a different reporter, whereby the pathway(s) affected by the test compounds can be elucidated.

10 114. The method of claim 109, wherein a reporter is operatively linked to a promoter that is transcriptionally regulated in response to DNA damage, and the test compounds are genotoxicants.

115. The method of claim 114, wherein the DNA damage is induced by apoptosis, necrosis or cell-cycle perturbations.

15 116. The method of claim 114, wherein unknown compounds are screened to assess whether they are genotoxicants.

117. The method of claim 114, wherein the promoter is a cytochrome P450-profiled promoter.

20 118. The method of claim 114, wherein the cell is in a transgenic animal and toxicity is assessed in the animal.

119. The method of claim 109, wherein:

the cell is a patient cell sample; the patient has a disease;

the regulatory region is one targeted by a drug or drug regimen;

and

25 the method assesses the effectiveness of a treatment for the disease for the particular patient.

120. The method of claim 119, wherein the cell is a tumor cell.

121. The method of claim 109, wherein the cell is a stem cell or a progenitor cell, whereby expression of the reporter is operatively linked to

-152-

a regulatory region expressed in the cells to thereby identify stem cells or progenitor cell.

122. The method of claim 109, wherein the cell is in an animal;
and the method comprises whole-body imaging to monitor expression of
5 the reporter in the animal.

123. A reporter *ACes* comprises one or a plurality of reporter constructs, wherein the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

Fig. 1 Generation of ACes for Platform Chromosome Engineering

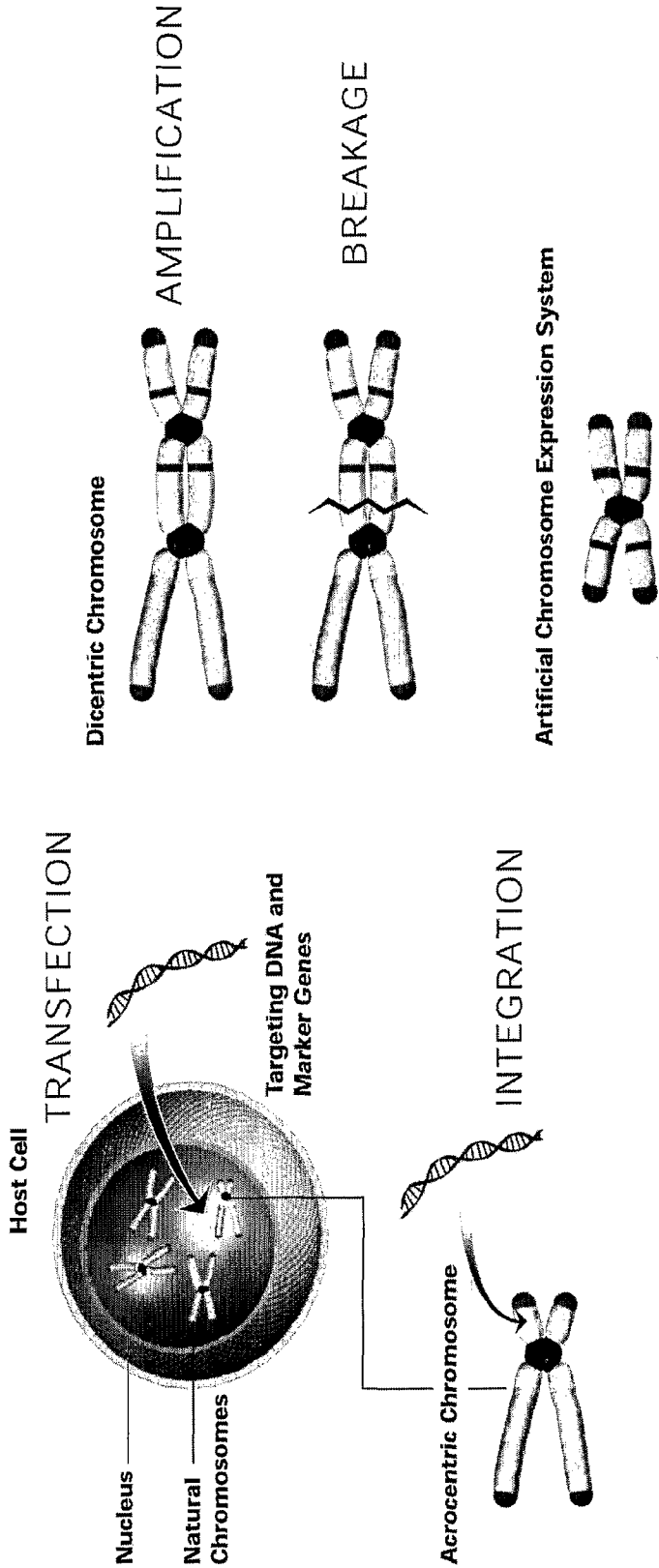


Fig. 2 Targeting plasmid pWEPuro9K

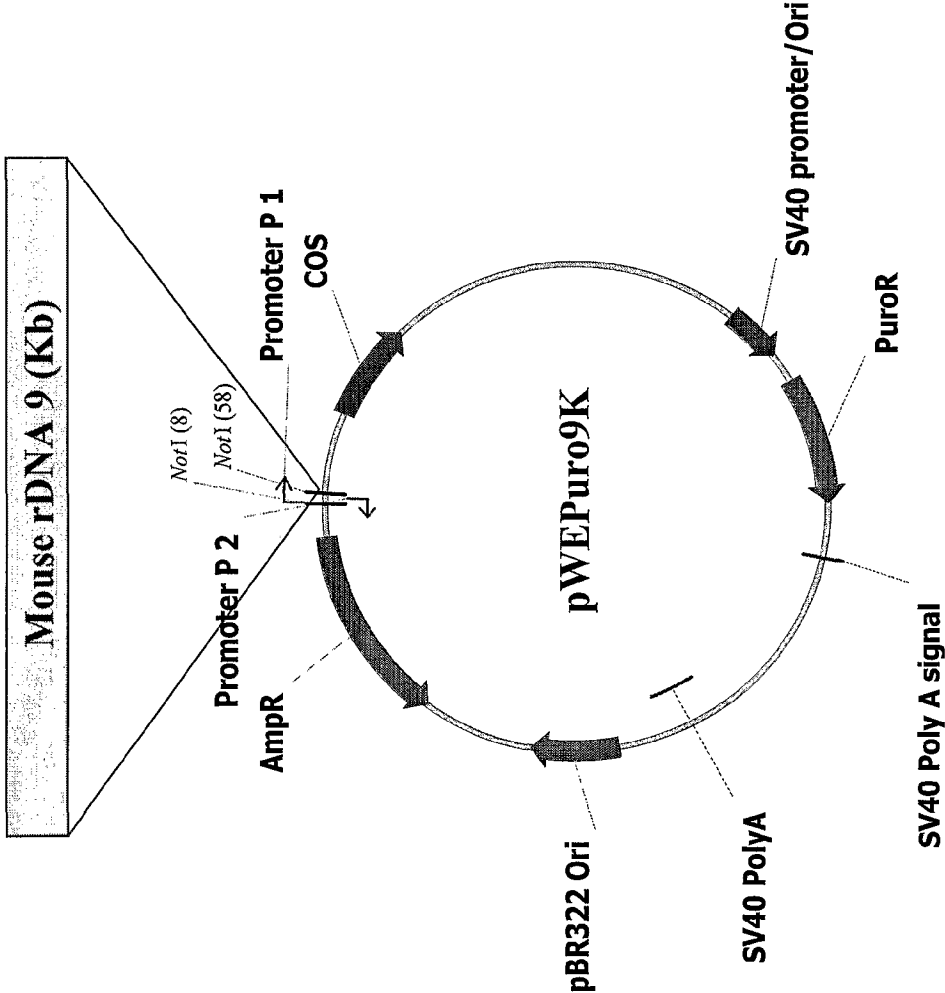


Fig. 3 Platform Chromosome with single recombination site

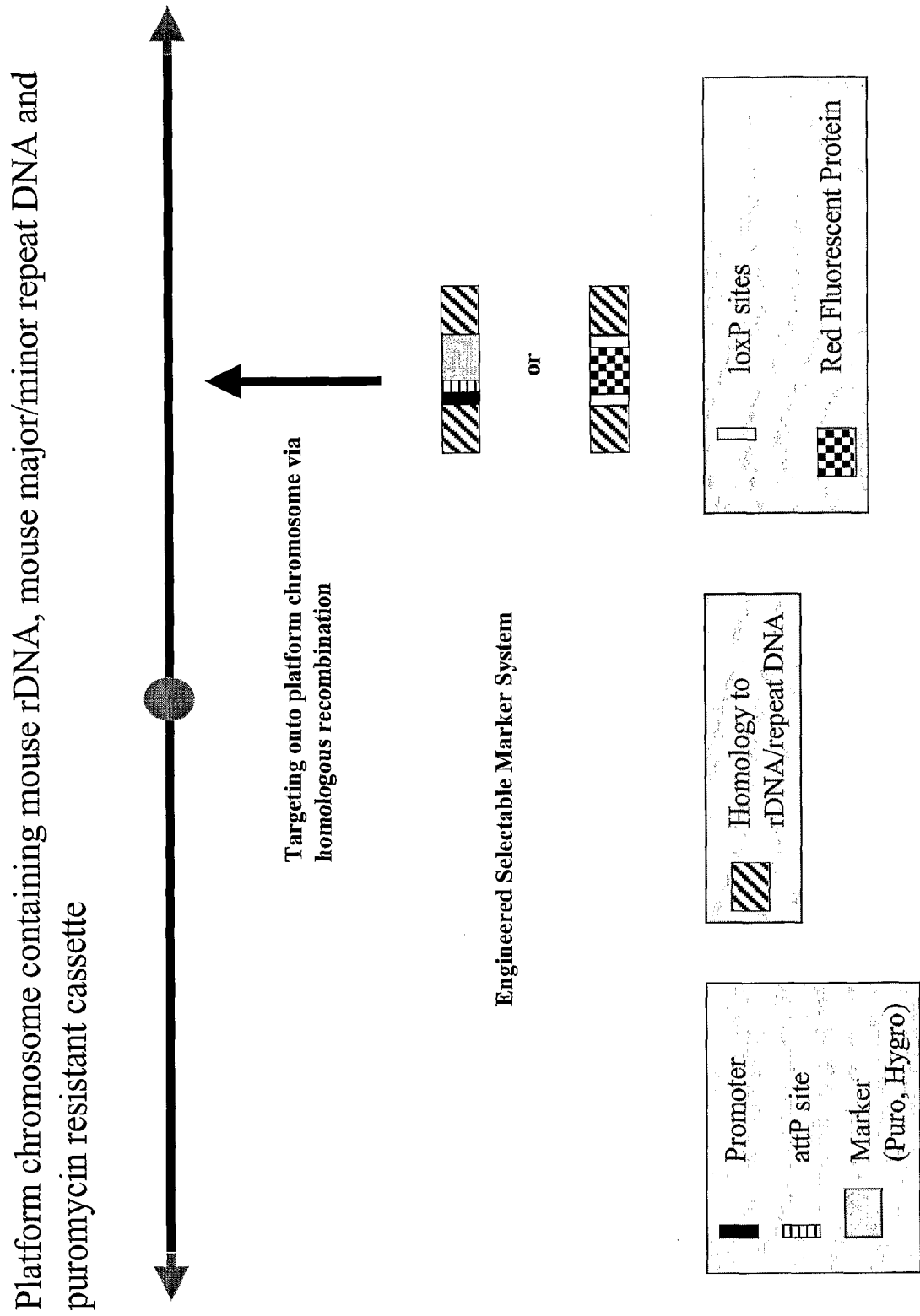


Fig. 4

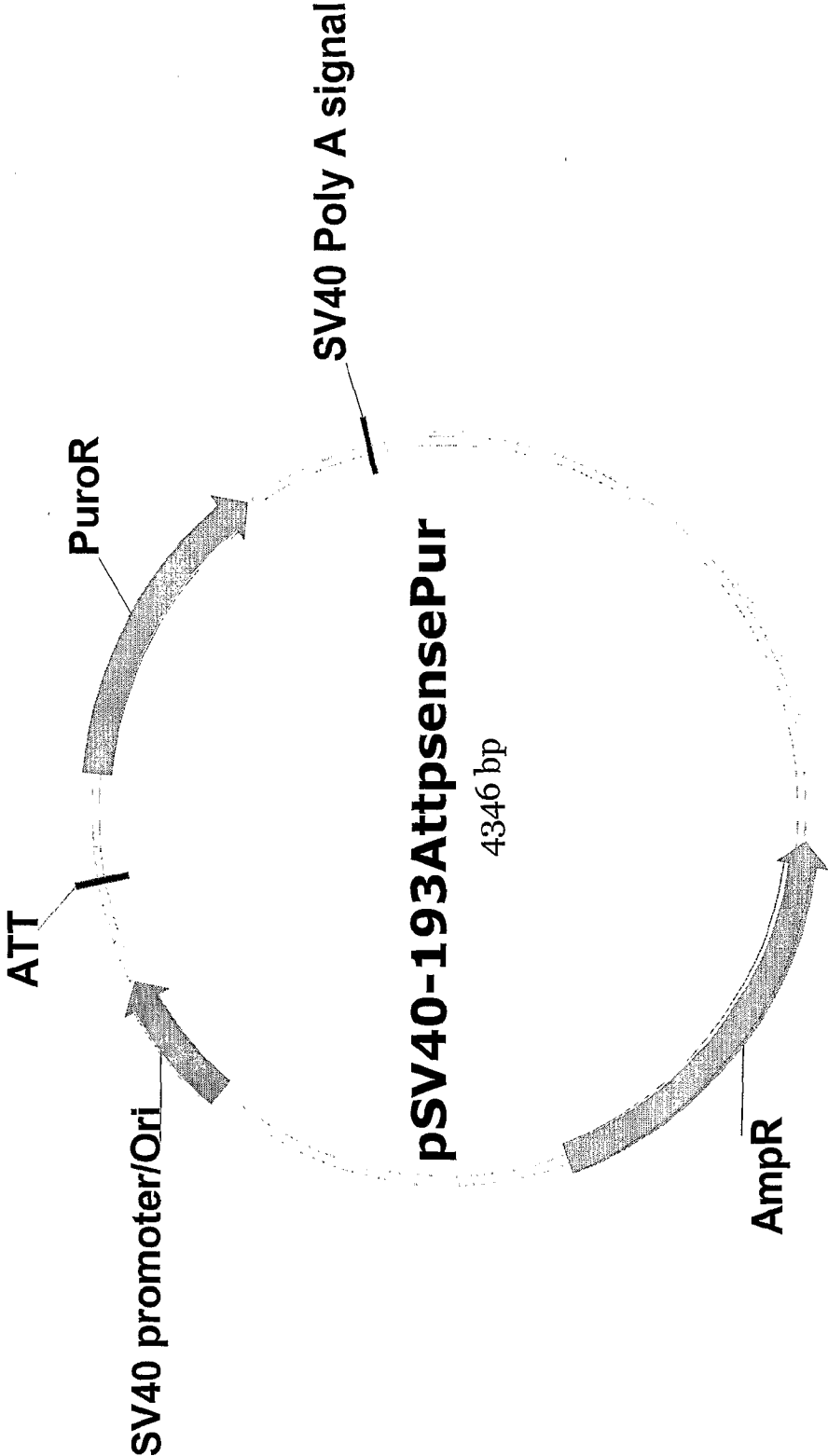


Fig. 5 Generation of chromosome based platform with multiple integration sites

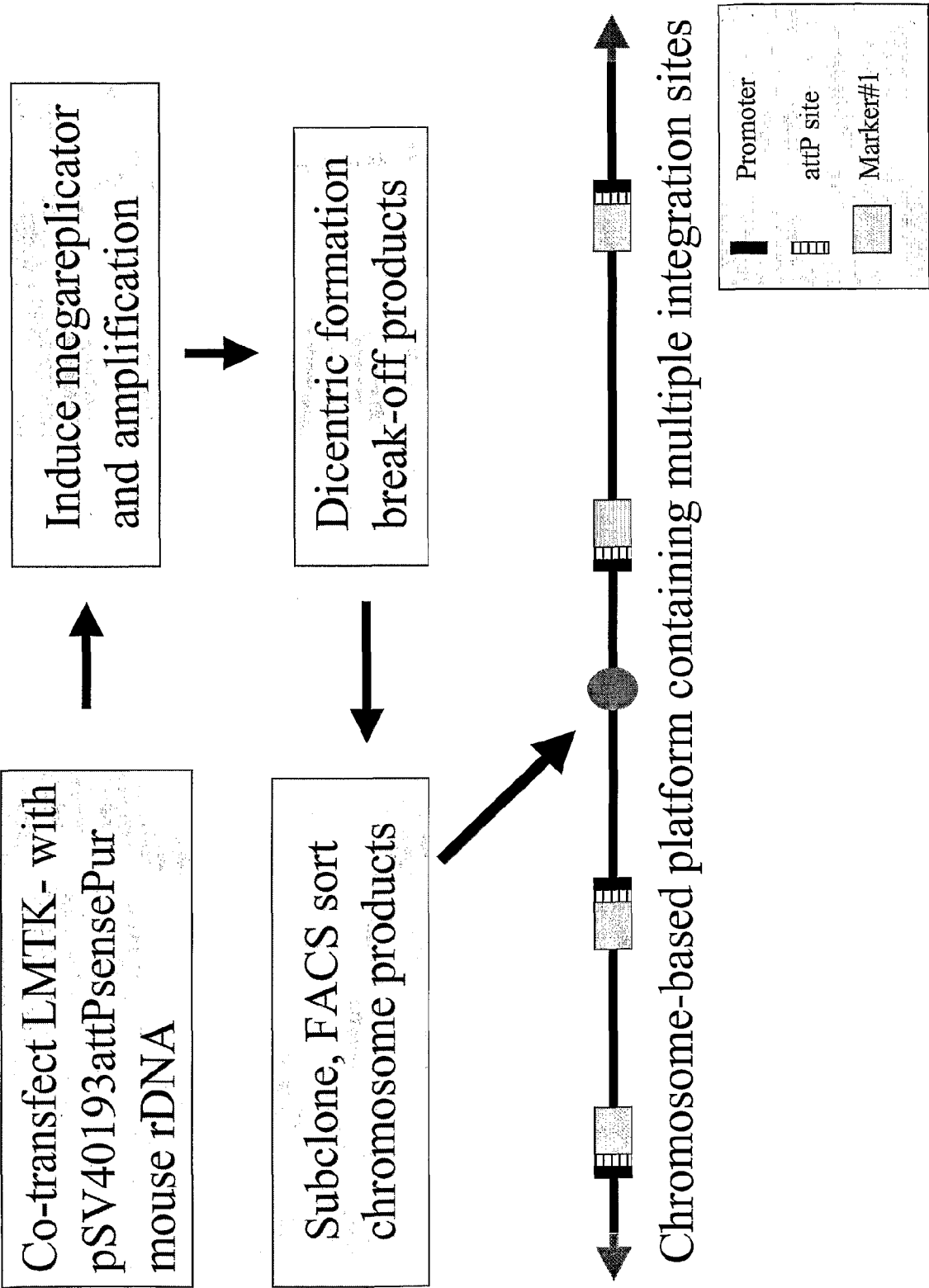


Fig. 6 λ integrase recombination



Core Region	
attP	CAGCTTTT <u>T</u> TATACTAA <u>G</u> TTG
attB	CTGCTTTT <u>T</u> TATACTAA <u>C</u> TTG
attL	CTGCTTTT <u>T</u> TATACTAA <u>G</u> TTG
attR	CAGCTTTT <u>T</u> TATACTAA <u>C</u> TTG

Fig. 7 λ INT recombination on artificial chromosome

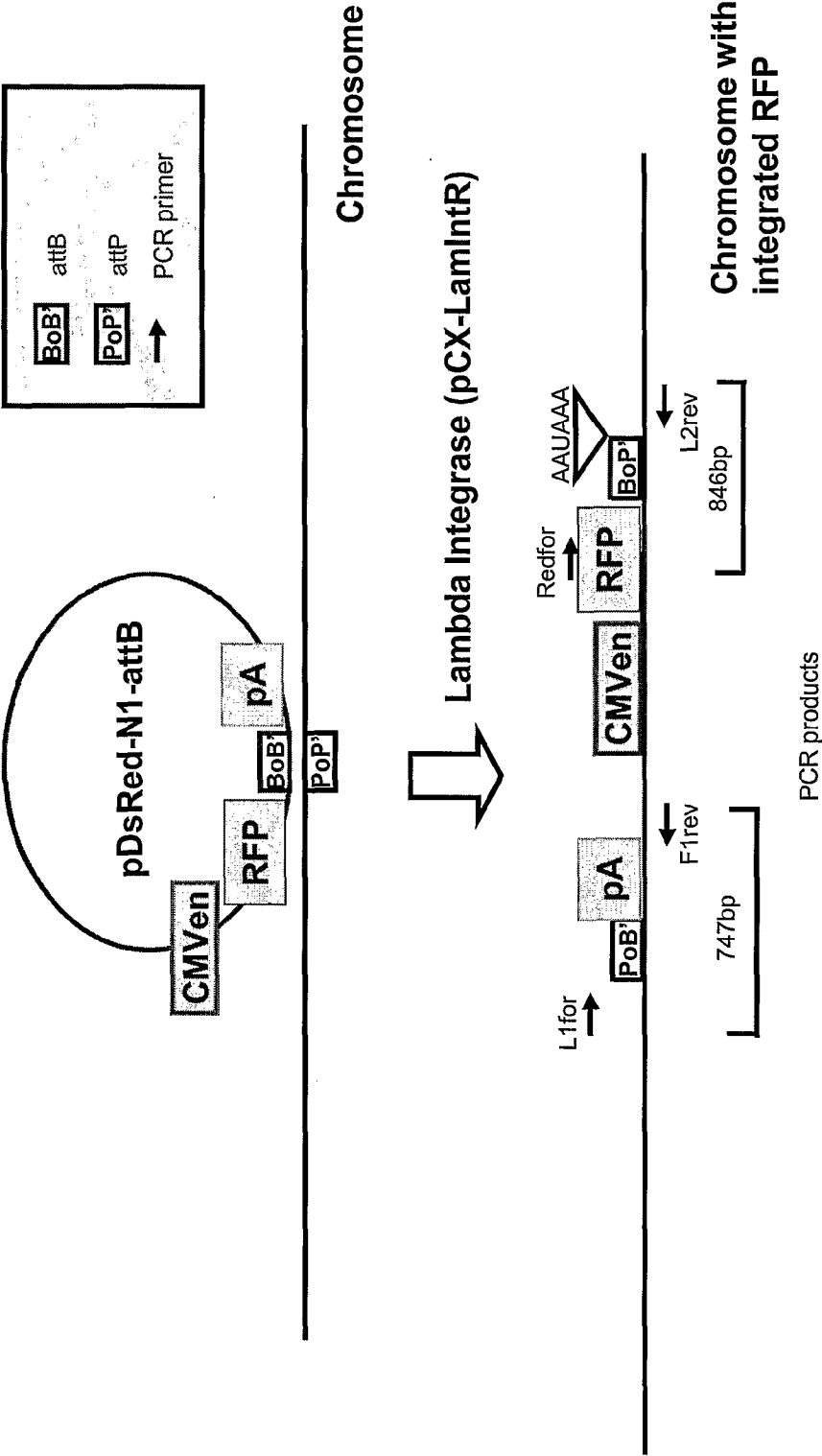
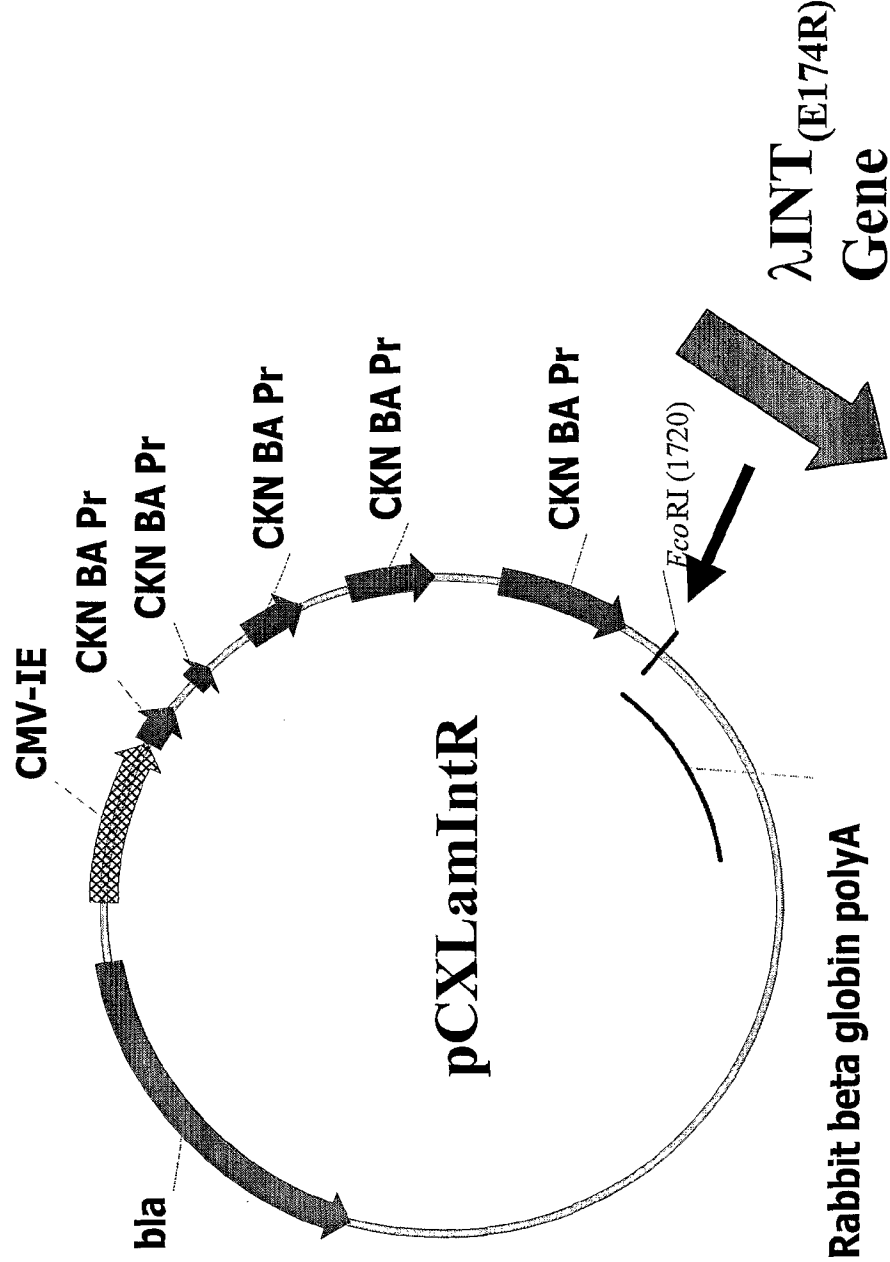


Fig. 8 pCX_{Lam}IntR integrase expression vector



9/15

Fig. 9 Platform Technology

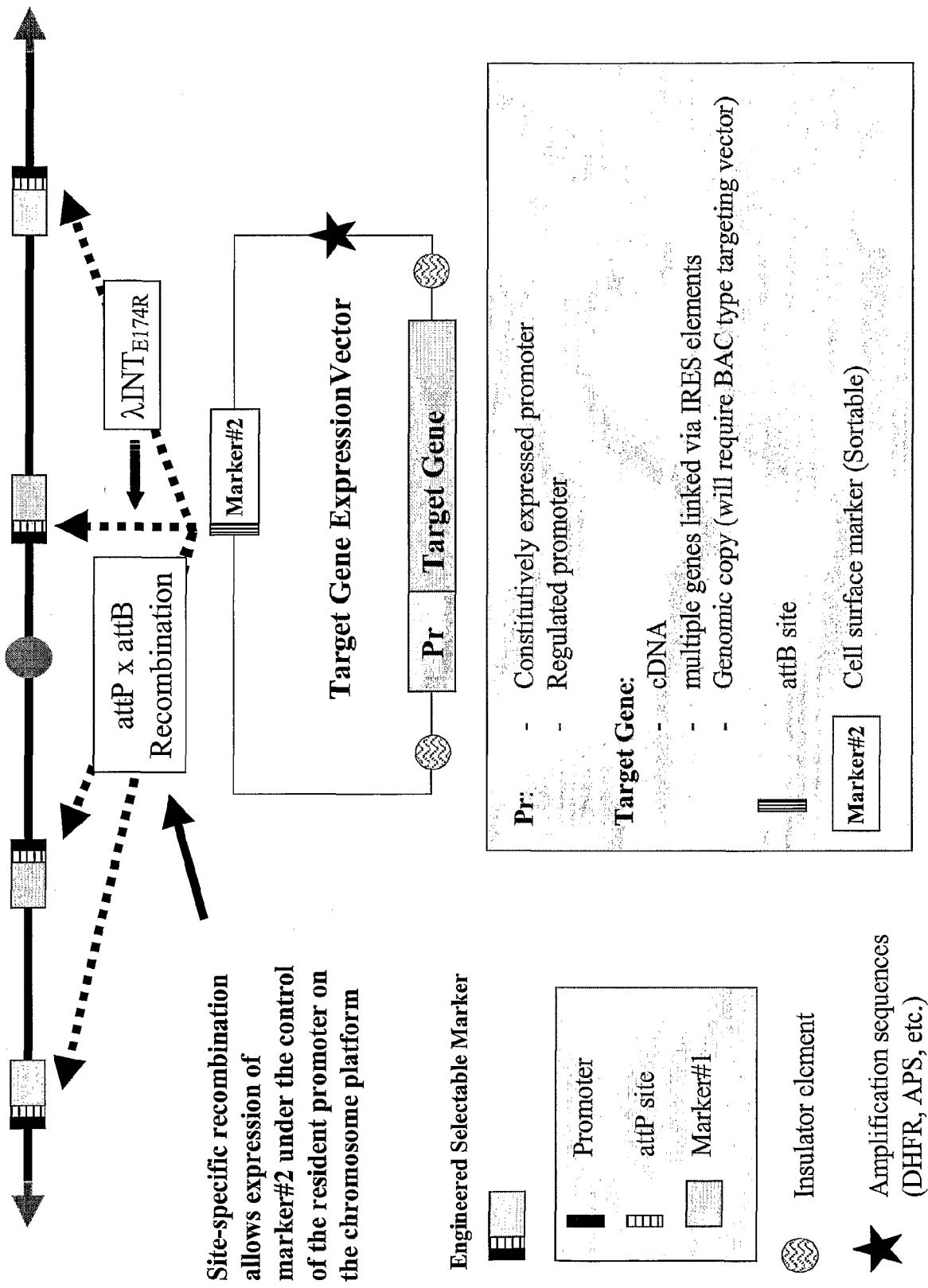


Fig. 10

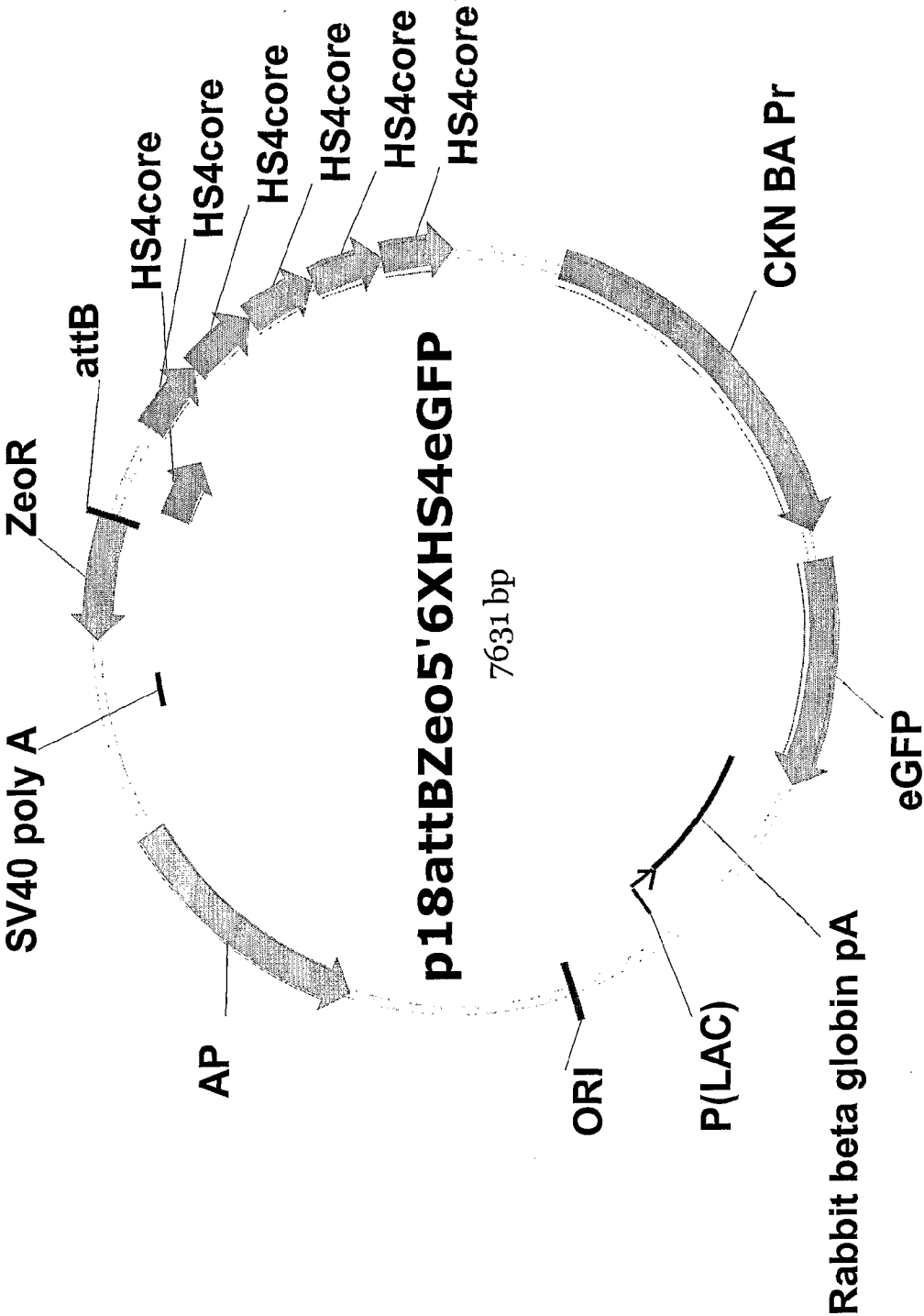


Fig. 11

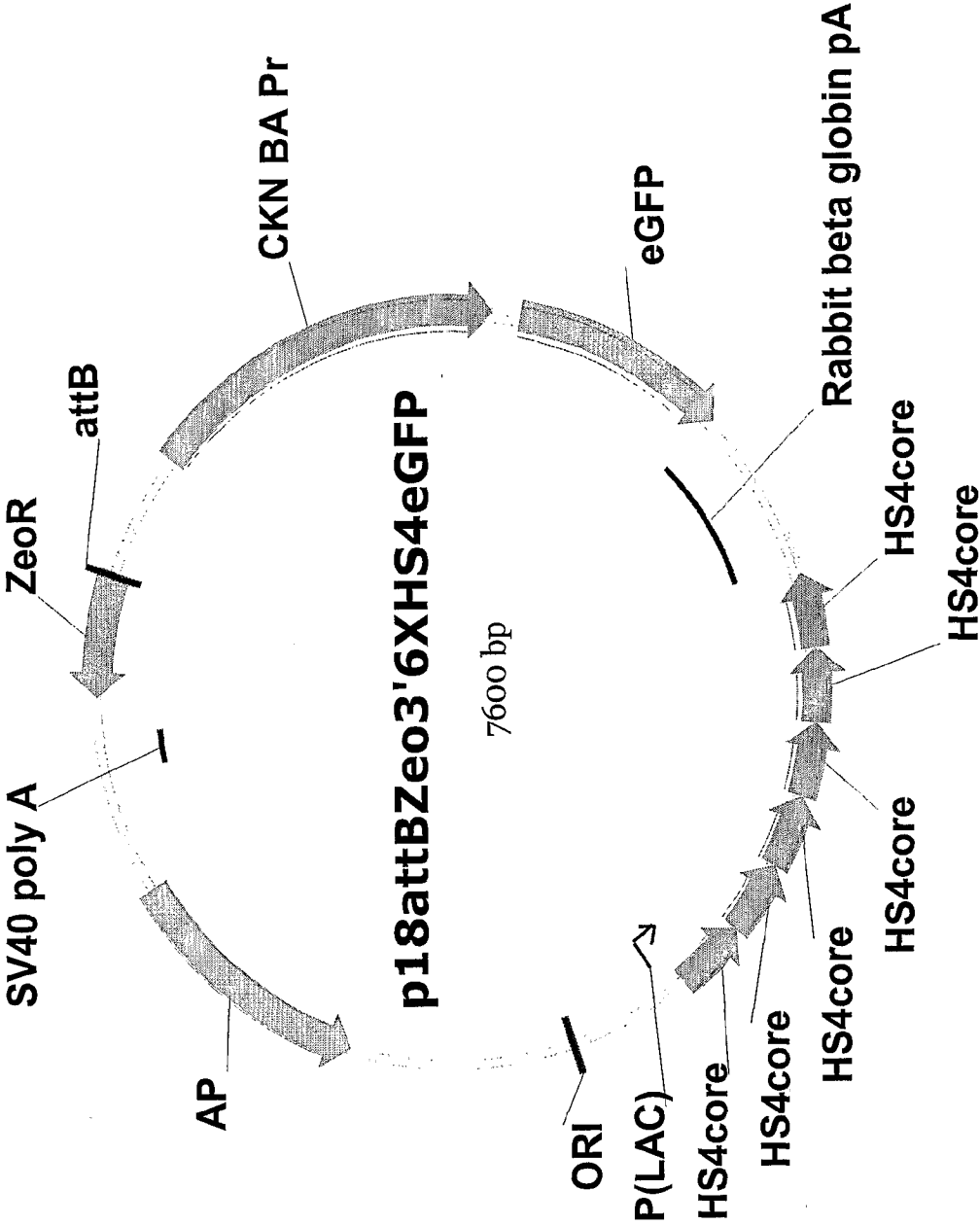


Fig. 12

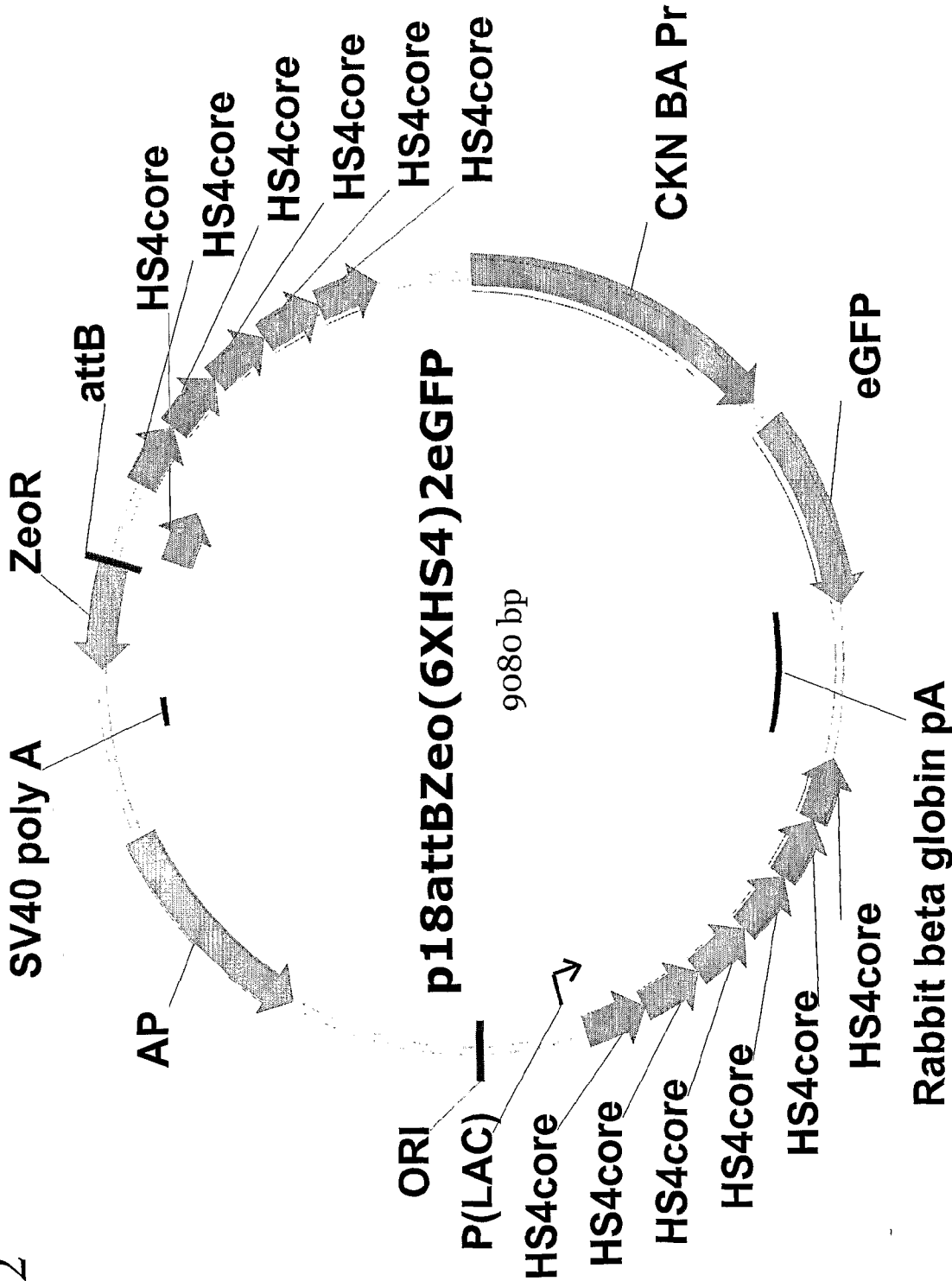
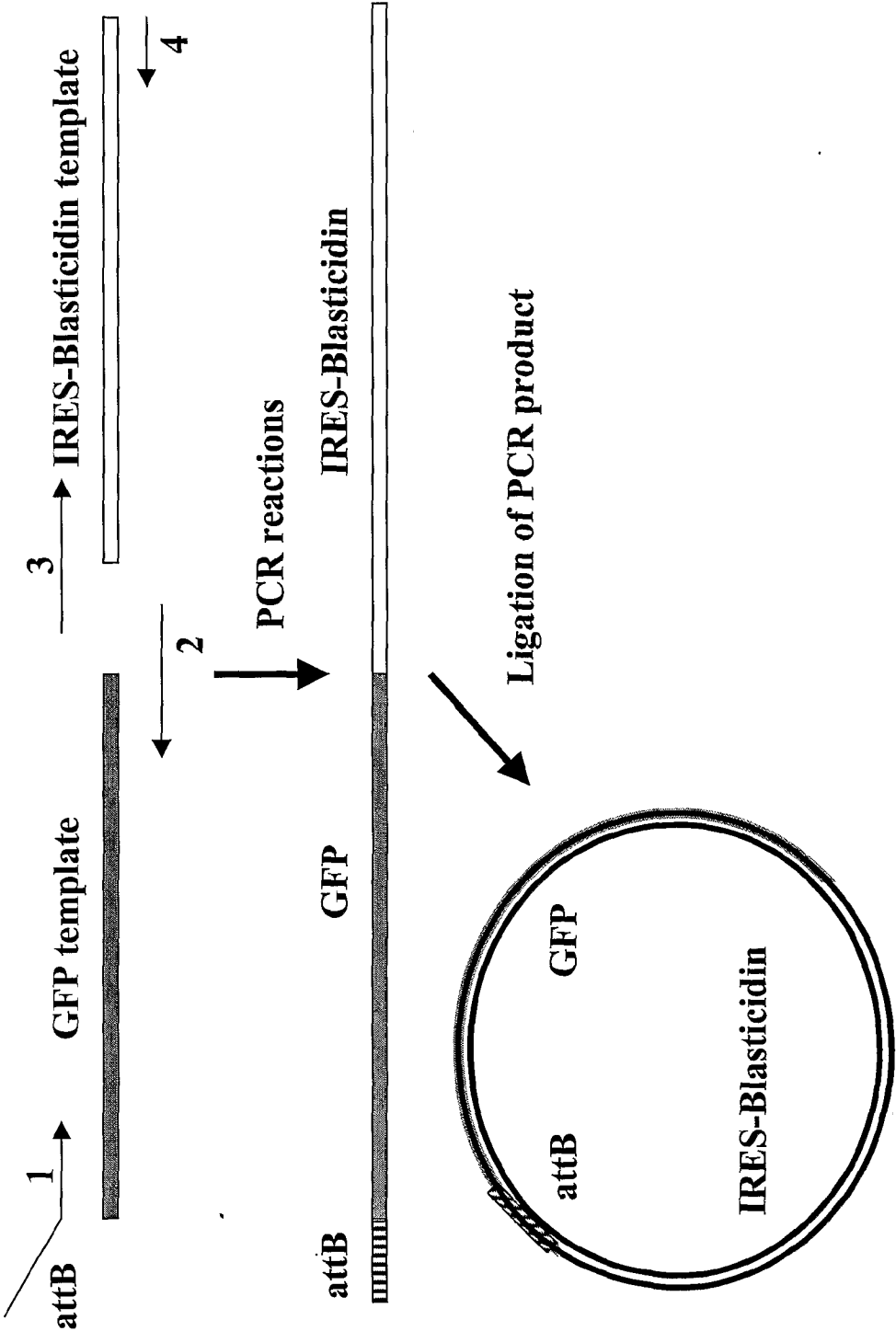
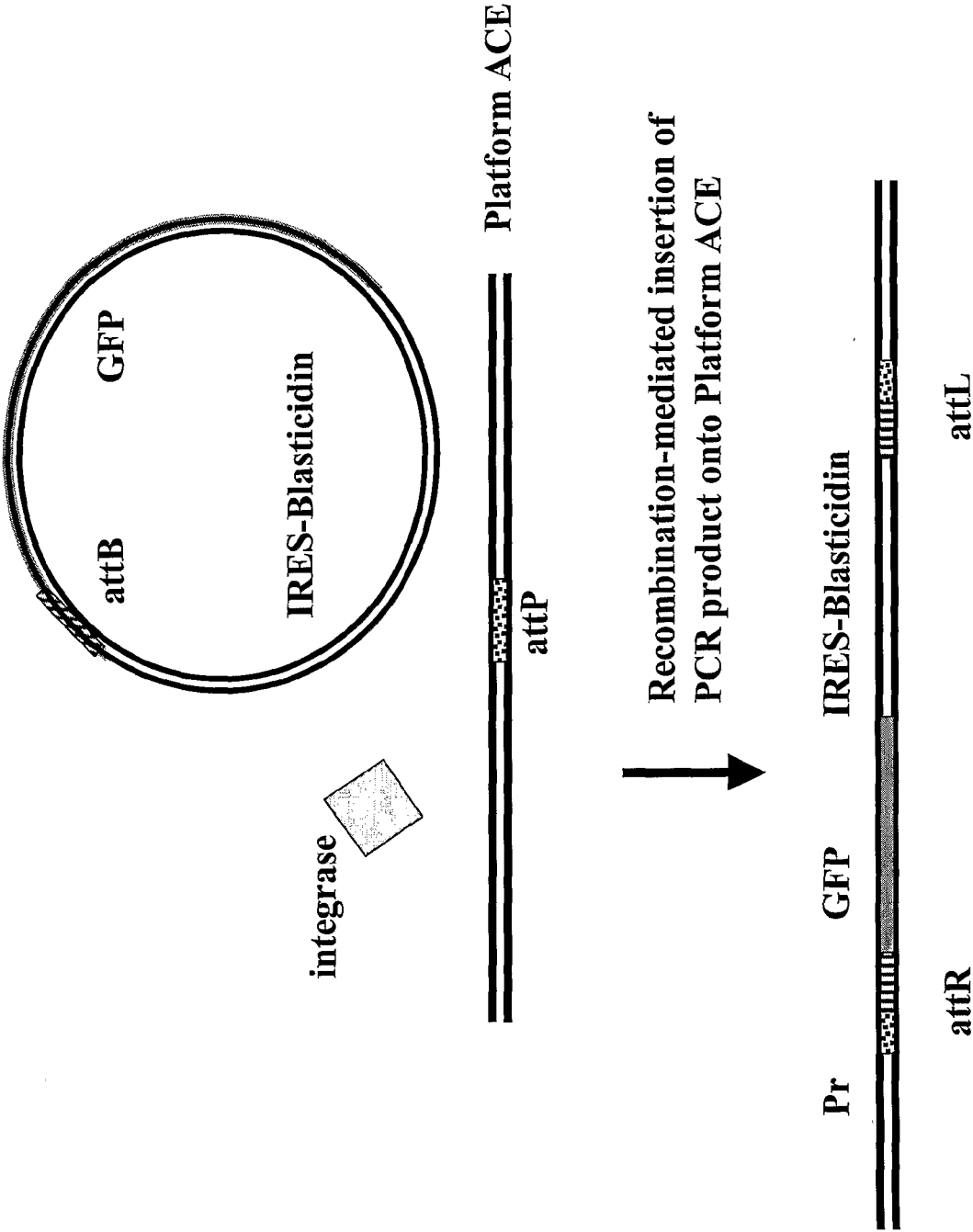


Fig. 13





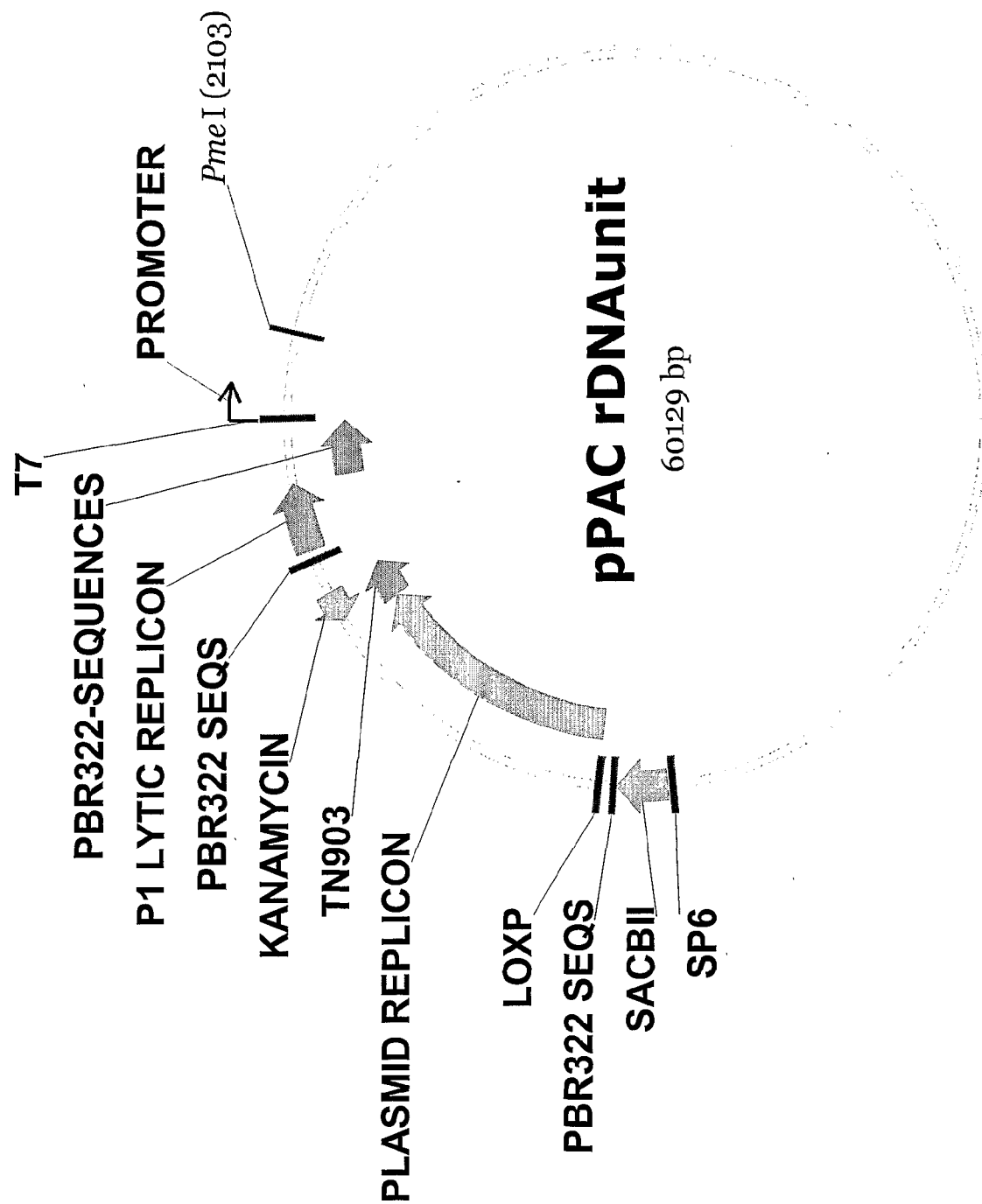


Fig. 15

-1-

SEQUENCE LISTING

<110> CHROMOS MOLECULAR SYSTEMS, INC.
Perkins, Edward
Perez, Carl
Lindenbaum, Michael
Greene, Amy
Leung, Josephine
Fleming, Elena
Stewart, Sandra
Shellard, Joan

<120> CHROMOSOME-BASED PLATFORMS

<130> 24601-420PC

<140> Not Yet Assigned
<141> Herewith

<150> 60/294,758
<151> 2001-05-30

<150> 60/366,891
<151> 2002-03-21

<160> 129

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer: attPUP

<400> 1
ccttgcgcta atgctctgtt acagg 25

<210> 2
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer: attPDWN

<400> 2
cagaggcagg gagggggaca aaattg 26

<210> 3
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer: Lamint 1

<400> 3
ttcgaattca tgggaagaag gcgaagtcac gagcg 35

<210> 4
<211> 34
<212> DNA
<213> Artificial Sequence

-2-

<220>
<223> Primer: Lamint 2

<400> 4
ttcgaattct tatttgattt caattttgtc ccac 34

<210> 5
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 5
cggacaatgc ggttggtgcgt 20

<210> 6
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 6
cgcgcagcaa aatctagagt aaggagatca agacttacgg ctgacg 46

<210> 7
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> LambdaINTER174rev

<400> 7
cgtcagccgt aagtcttgat ctcttactc tagattttgc tgcgcg 46

<210> 8
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> attB1

<400> 8
tgaagcctgc ttttttatac taacttgagc gaa 33

<210> 9
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> attB2

<400> 9
ttcgctcaag ttagtataaa aaagcaggct tca 33

<210> 10
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

-3-

<223> Primer: attPdwn2

<400> 10
tcttctcggg cataagtcgg acacc 25

<210> 11
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:CMVen

<400> 11
ctcacgggga tttccaagtc tccac 25

<210> 12
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:attPdwn

<400> 12
cagaggcagg gagtgggaca aaattg 26

<210> 13
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:CMVEN2

<400> 13
caactccgcc ccattgacgc aaatg 25

<210> 14
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:L1

<400> 14
agtatcgccg aacgattagc tcttca 26

<210> 15
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:F1 rev

<400> 15
gccgatttcg gcctattggt taaa 24

<210> 16
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:RED

-4-

<400> 16
ccgccgacat ccccgactac aagaa 25

<210> 17
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:L2rev

<400> 17
ttccttcgaa ggggatccgc ctacc 25

<210> 18
<211> 22118
<212> DNA
<213> Mus musculus

<300>
<308> GenBank X82564
<309> 1996-04-09

<400> 18
gaattccctc atccctaata cagattgggtg gaataaacttg gtatagatgt ttgtgcatta 60
aaaaccctgt aggatcttca ctctaggtca ctgttcagca ctggaacctg aattgtggcc 120
ctgagtata ggtcctggga catatgcagt tctgcacaga cagacagaca gacagacaga 180
cagacagaca gacagacgtt acaaacaaac acgttgagcc gtgtgccaac acacacacaa 240
acaccactct ggccataatt attgaggacg ttgatttatt attctgtgtt tgtgagtctg 300
tctgtctgtc tgtctgtctg tctgtctgtc tatcaaacca aaagaaacca aacaattatg 360
cctgcctgcc tgcctgcctg cctacacaga gaaatgattt cttcaatcaa tctaaaacga 420
cctcctaagt ttgccttttt tctctttctt tatctttttc ttttttcttt tcttcttctt 480
tcttctcttc cttccttctt tcttcttctt ctttcttctt ttcttcttctt cttactttct 540
ttcttctctt cttacattta ttcttttcat acatagtctt ttagtgtaag catccctgac 600
tgtcttgaag acactttgta ggctcaatc ctgtaagagc cttcctctgc ttttcaaagt 660
ctggcaatgaa tgttgtacct cactatgacc agcttagtct tcaagtctga gttactggaa 720
aggagtcca agaagactgg ttatatTTTT catttattat tgcattttta ttaaaattta 780
atttcaccaa aagaatttag actgaccaat tcagagtctg ccgtttaaaa gcataaggaa 840
aaagtaggag aaaadactga ggctgtctgt ggatggtcga ggctgcttta gggagcctcg 900
tcaccattct gcacttgcaa accgggccac tagaaccogg tgaagggaga aaccaaagcg 960
acctggaaac aataggtcac atgaaggcca gccacctcca tcttggtgtg cgggagttca 1020
gttagcagac aagatggctg ccatgcacat gttgtctttc agcttggtga ggtcaaagta 1080
caaccgagtc acagaacaag gaagtatata cagtgagttc caggtcagcc agagtttaca 1140
cagagaaac acatcttgaa aaaaaacaaa aaataaata aataaatata atttaaaaaa 1200
ttaaaaatag cgggcatgtg tggcgcatgt ctttaatccc agctctcttc aggcagagat 1260
gggaggattt ctgagtttga ggccagcctg gtctgcaaag tgagttccag gacagtcagg 1320
gctatacaga gaaaccctgt cttgaaaact aaactaaatt aaactaaact aaactaaaaa 1380
aatataaaat aaaaatttta aagaatttta aaaaactaca gaaatcaaac ataagcccac 1440
gagatggcaa gtaactgcaa tcatagcaga aatattatac acacacacac acacagactc 1500
tgtcataaaa tccaatgtgc cttcatgatg atcaaaatttc gatagtcagt aatactagaa 1560
gaatcatatg tctgaaaata aaagccagaa ccttttctgc ttttgttttc ttttgcccca 1620
agataggggt tctctcagtg tatccctggc atccctgcct ggaacttcct ttgtaggttt 1680
ggtagcctca aactcagaga ggtcctctct gcctgcctgc ctgcctgcct gcctgcctgc 1740
ctgcctgcct gcctgcctca cttcttctgc cttcttctct ttcttctctt tttcttcttt 1800
tttatttctt tctcttctct tcttcttctt ctttcttctt ctttcttctt tttcttcttt 1860
ctttcttctt tttcttattc attagttttc aatgtaagtg tgtgtttgtg ctctatctgc 1920
tgccctatagg cctgcttgcc aggagagggc aacagaacct aggagaaacc accatgcagc 1980
tcctgagaaat aagtgaaaaa acaacaaaaa aaggaaattc taatcacata gaatgtagat 2040
atatgccgag cctgtcagag tgctttttta ggcttagtgt aagtaatgaa aattgttgtg 2100
tgtcttttat ccaaacacag aagagaggtg gctcggcctg catgtctgtt gtctgcattg 2160
agaccaggct ggccttgaac acattaatct gtctgcctct gcttccctaa tgctgcgatt 2220
aaaggcatgt gccaccactg cccggactga tttcttcttt tttttttttt tggaaaatac 2280
ctttcttctt ttttctctct ctcttctctt cttccttctt ttcttctctt tcttttcttt 2340
tttcttcttt tttttttttt ttttttttaa aatttgccca aggttaaagg tgtgctccac 2400
aattgcctca gctctgctct aattctcttt aaaaaaaaac aaacaaaaaa aaaacaaaaa 2460
cagtatgtat gtatgtatat ttagaagaaa tactaatcca ttaataactc ttttttctta 2520
aaattcatgt cattcttgtt ccacaaagtg agttccagga cttaccagag aaacctgtg 2580

-5-

ttcaaatttc	tgtgttcaag	gtcacccctgg	cttacaaagt	gagttccaag	tccgataggg	2640
ctacacagaa	aaaccatata	tcagaaaaaa	aaaaagtccc	aaacacacac	acacacacac	2700
acacacacac	acacacacac	acacacacac	acacacacac	cgcgccgcgg	cgatgagggg	2760
aagtctgtcc	taaaataaat	atTTTTctgg	ccaaagtga	agcaaatacac	tatgaagagg	2820
tactcctaga	aaaaataaat	acaaacgggc	TTTTtaata	ttccagcact	gttttaattt	2880
aactctgaat	ttagtcttgg	aaaagggggc	gggtgtgggt	gagtgagggc	gagcgagcag	2940
acgggccccg	gggccccgga	gtggccccgc	gcggtggcag	cgagcaccag	aaaacaacaa	3000
acccccaggc	gtagagtgtt	ttaaaaatga	gacctaaatg	tggtggaacg	gaggtcgccg	3060
ccaccctcct	cttccactgc	ttagatgctc	ccttccccct	actgtgctcc	cttccccctaa	3120
ctgtgcctaa	ctgtgcctgt	tccctcacc	cgctgattcg	ccagcgacgt	actttgactt	3180
caagaacgat	tttgccctgt	ttcacgcctc	cctgtcatat	tttcgttttt	gggtgccccg	3240
gtctagcccg	ttcgctatgt	tcgggccccg	cgatggggac	cgtttgtgcc	actcgggaga	3300
agtgggtggg	gggtacgctg	ctccgtcgtg	cgctgctgag	tgccggaacc	tgagctcggg	3360
agaccctccg	gagagacaga	atgagtgaat	gaatgtggcg	gcgctgacag	gatctgtatt	3420
gggtttgtatg	gttgatcgag	accattgtcg	ggcgacacct	agtggtgaca	agtttcggga	3480
acgctccagg	cctctcaggt	tggtgacaca	ggagaggga	gtgctgtgg	tgaggcgacc	3540
agggcgacag	gaggccgggc	aagcaggcgg	gagcgtctcg	gagatgggtg	cggttttaag	3600
gacggtctct	aaacaaggag	tcgtacagg	agatggccaa	agcagaccga	gttgctgtac	3660
gcccttttgg	gaaaaatgct	agggttgggt	gcaacgttac	taggtcgacc	agaaggctta	3720
agtcctaccc	ccccccccct	tttttttttt	ttcctccag	aagccctctc	ttgtccccct	3780
caccgggggg	acogtacatc	tgaggccgag	aggacgcgat	gggccccggt	tccaagccgg	3840
tggtggctcg	ccagctggcg	cttcgggtct	tttttttttt	tttttttttt	ttttcccca	3900
gaagccttgt	ctgtcgctgt	caccgggggc	gctgtacttc	tgaggccgag	aggacgcgat	3960
gggccccggc	ttccaagccg	gtgtggctcg	gcagctgga	gcttcgggtc	tttttttttt	4020
tttttttttt	ttttttttct	cagaagcctt	gtctgtcgct	gtcaccgggg	gcgctgtact	4080
tctgagcccg	agaggacgcg	atgggtcgcc	ttccaagccg	atgtggcggg	gccagctgga	4140
gcttcggggt	tttttttttc	ctccagaagc	cctctcttgt	ccccgtcacc	gggggcgctg	4200
tacttctgag	gcccagagga	cgatgatggc	ccgggttcca	ggcggatgtc	gcccggtcag	4260
ctggagcttt	ggatcttttt	tttttttttt	cctccagaag	ccctctcttg	tccccgtcac	4320
cggggggcacc	ttacatctga	gggcgagagg	acgtgatggg	tccggcttcc	aagccgatgt	4380
ggcggggcca	ctcgagctt	cgggtttttt	tttttctct	cagaagccct	ctctgtctcc	4440
cgtcacccgg	ggcgctgtac	ttctgaggcc	gagaggacgt	gatgggcccg	ggttccaggc	4500
ggatgtcgcc	cggtcagctg	gagctttgga	tcattttttt	ttttccctcc	agaagccctc	4560
tcttgtcccc	gtcaccgggg	gcaccgtaca	tctgaggccg	agaggacacg	atgggcctgt	4620
cttccaagcc	gatgtggccc	ggccagctgg	agcttcgggt	cttttttttt	ttttttcttc	4680
cagaagcctt	gtctgtcgct	gtcaccggcg	gcgctgtact	tctgaggccg	agaggacgcg	4740
atgggccccg	cttccaagcc	gggtgtggct	ggccagctgg	agcttcgggt	cttttttttt	4800
tttttttttt	ttcctccaga	aaccttgtct	gtcgtgtca	cccggggcgc	ttgtacttct	4860
gatgccgaga	ggacgcgatg	ggcccgctct	ccaggccgat	gtggccccgt	cagctggagc	4920
tttggaatct	tttttttttt	ttttctcca	gaagccctct	cttgtccccg	tcaccggggg	4980
caccttacat	ctgaggccta	gaggacacga	tgggccccgg	ttccaggccg	atgtggcccc	5040
gtcagctgga	gctttggatc	tttttttttt	ttttcttcca	gaagccctct	tgtccccgtc	5100
accgggtggc	ctgtacatct	gaggcgagga	ggacattatg	ggccccgctt	ccaatccgat	5160
gtggccccgt	cagctggagc	tttggaatct	attttttttt	taattttttt	ttccagaagc	5220
cctcttgtcc	ctgtcaccgg	tggcacggta	catctgaggc	cgagaggaca	ttatgggccc	5280
ggccttcagg	ccgatgtggc	ccggtcagct	ggagctttgg	atcttttttt	ttttttttct	5340
tttttccctc	agaagccctc	tctgtccctg	tcaccggggg	ccctgtacgt	ctgaggccga	5400
gggaaagcta	tgggcgcggg	tttctttcat	tgacctgtcg	gtcttatcag	ttctccgggt	5460
tgtcagggtc	gaccagttgt	tcccttgagg	tccggttctt	ttcgttatgg	gggtcatttt	5520
gggccaacct	cccaggatat	acttccaggc	gtcgttgtct	gctgtcact	ttcctccctg	5580
tctcttttat	gcttgtgatc	ttttctatct	gttccctatt	gacctggaga	taggtactga	5640
cacgctgtcc	tttccctatt	aacactaaag	gacactataa	agagaccctt	tcgatttaag	5700
gctgttttgc	ttgtccagcc	tattcttttt	actggcttgg	gtctgtcgcg	gtgcctgaag	5760
ctgtccccgc	gccacgcttc	ctgctttccc	gggcttgcgt	cttgctgtgt	cttgctgtgg	5820
gcagctttgt	acaactgggc	gctgtgactt	tgtcgtgtgt	cagacgtttt	tcccgatttc	5880
cccaggggtg	cgttgtcaca	cctgtcccgg	ttggaatggg	ggagccagct	gtgggtgagg	5940
gccaccttat	ttcggtccac	tttttttttt	tttttttctc	ttggagtcac	gaacctccgc	6000
tcttttctct	tcccggctct	tcttccacat	gcctccccgag	tgcattttct	tttgtttttt	6060
ttcttttttt	tttttttttt	ttggggagg	ggagagtcac	gagtagcttc	ctcctgtctg	6120
tggtgtccaa	gtgttcatgc	cacgtgcctc	ccgagtgcac	ttttttttgt	ggcagtcgct	6180
cgttgtgttc	tcttgttctg	tgtctgccc	tatcagtaac	tgtcttgccc	cgcgtgtaag	6240
acattcctat	ctcgcttggt	tctcccgat	gcgctgcgtt	gctcactctt	agatcgatgt	6300
gggtgctccg	agttctcttc	gggcccaggg	caagccgcgc	caggccgagg	acggacattc	6360
atggcgatg	gcggccgctc	ttctcgttct	gccagcgggc	cctcgtctct	ccaccaccatc	6420
cgtctgcccg	tggtgtgtgg	aaggcagggg	tgcggctctc	cggcccgacg	ctgcccccg	6480
cgcacttttc	tcagtgggtc	gcgtgggtct	tgtggatgtg	tgaggcgccc	ggttgtgccc	6540
tcacgtgttt	cactttgggt	gtgtctcgct	tgaccatgtt	cccagagtcg	gtggatgtgg	6600

-6-

cgggtggcgt	tgcataccct	tcccgtctgg	tgtgtgcacg	cgetgtttct	tgtaaagcgtc	6660
gaggtgctcc	tggagcgttc	caggtttgtc	tcctaggtgc	ctgcttctga	gctgggtggtg	6720
gcgctcccca	ttccctgggtg	tgcctccggt	gctccgtgtg	gctgtgtgcc	ttcccgtttg	6780
tgtctgagaa	gcccgtgaga	gggggggtcga	ggagagaagg	aggggcaaga	ccccctttct	6840
tcgtcgggtg	aggcgccac	cccgcgacta	gtacgcctgt	gcgtaggggt	ggtgctgagc	6900
ggtcgcggct	gggggttgaa	agtttctcga	gagactcatt	gctttcccg	ggggagcttt	6960
gagagccctg	gctttcgggg	gggaccggtt	gcagggtctc	ccctgtccgc	ggatgctcag	7020
aatgcccttg	gaagagaacc	ttcctgttgc	cgcagacccc	cccgcgcggt	cgcccgcggtg	7080
ttggtcttct	ggtttccctg	tgtgctcgtc	gcctgcctcc	tctctcggtg	gccgggggtc	7140
gtcgggggtt	tgggtccgctc	ccgcctcag	tgagaaagtt	tccttctcta	gctatcttcc	7200
ggaaagggtg	cgggcttctt	acggtctcga	ggggtctctc	ccgaatgggtc	ccctggaggg	7260
ctcgccccc	gaccgcctcc	cgcgcgcgca	gcgtttgctc	tctcgtctac	cgcggcccg	7320
ggcctccccg	ctccgagttc	ggggagggat	cacgcggggc	agagcctgtc	tgtcgtcctg	7380
ccgttgctgc	ggagcatgtg	gctcggcttg	tgtgggtggg	ggctggggag	agggctccgt	7440
gcacaccccc	gcgtgcgcgt	actttcctcc	cctcctgagg	gcccgcgtgc	ggacgggggtg	7500
tgggtaggcg	acgggtgggct	cccgggtccc	caccgctctt	cccggtgcctc	accggtgcct	7560
tcgctcgctg	cggtccctct	cgctcgcgct	cacgactttg	gocgctcccg	cgacggcgcg	7620
ctgcgcgcg	cgtggtgcgt	gctgtgtgct	tctcgggtcg	tgtggttgtg	tcgcctcgcc	7680
cccccttcc	cgcggcagcg	ttcccacggc	tggcgaaatc	gcgggagtc	tccttcccc	7740
cctcgggggtc	gagagggtcc	gtgtctggcg	ttgattgatc	tcgctctcgg	ggacggggacc	7800
gttctgtggg	agaacggctg	ttggaccgct	cggcgcgac	gtcggacgtg	gggacccact	7860
gcgctcgggg	gggtctcgtc	ggtaggcac	gggtgtcgtg	catcggtctc	tctctcgtgt	7920
cgggtgcgcc	tctcgggct	cccggggggc	cgtcgtgttt	cgggtcggct	cggcgctgca	7980
ggtgtggtgg	gactgctcag	gggagtggtg	cagtgtgatt	cccgcgggtt	ttgcctcgcg	8040
tgccttgacc	ggtccgacgc	ccgagcggtc	tctcggctcc	ttgtgaggac	ccccctcccg	8100
gaggggcccc	tttcggccgc	ccttgccgct	gtcgccggcc	tctgtctcgc	tgtgtcgttc	8160
ccccctcccc	gctcgcgcga	gccgggtctt	tttctctct	ccccccctct	cctctgactg	8220
accggtggcc	gtgctgtcgg	accccccgca	tgggggcggc	cgggcacgta	cgcgtccggg	8280
cggtcacccg	ggtcttgggg	gggggcccag	gggtaagaaa	gtcggctcgg	cgggcggggag	8340
gagctgtggt	ttggagggcg	tcccggcccc	gcggccgtgg	cgggtgtctt	cgcgggtctt	8400
gagagggctg	cgtgcgaggg	gaaaagggtg	ccccgcgagg	gcaaagggaa	agaggctagc	8460
agtggctcatt	gtcccgcagg	tgtggtggtc	tgttgccga	ggtgcgtctg	gggggctcgt	8520
ccggccctgt	cgtccgtcgg	gaaggcgcgt	ggtggggcct	gccggagtgc	cgagggtggg	8580
accctggcgg	tgggattaac	cccgcgcgcg	tgtcccgggt	tggcggtggg	ggctccgggtc	8640
gatgtctacc	tccctctccc	cgaggctcca	ggccttctcc	gcgcgggctc	tcggccctcc	8700
cctcgttcc	ccctctcgcg	gggttcaagt	cgtcgtcga	ccctccctcc	tccgtccttc	8760
catctctcgc	gcaatggcgc	cgcctcagtt	cacggtgggt	tcgtcctccg	cctccgcttc	8820
tcgcgggggg	ctggccgctg	tccggtctct	cctgcccgac	cccgttggc	gtgggtcttct	8880
ctcgcgggtc	tcgcggactc	ctggcttcgc	ccggagggtc	agggggcttc	cgggttcccc	8940
gacgttgcgc	gctcgtctg	tgtgcttggg	gggggcgcgc	tgcggcctcc	gcccgcctcg	9000
gagccccctg	cgcacccgcc	gggtgtcggt	ttcgcgcgc	ggtcagttgg	gcccgtggcg	9060
tgtgtcgcgt	cgggagcgtg	tccgcctcgc	ggcggtcaga	cgcgggtgtc	gcccgggtcc	9120
gacgggtggc	ctatccaggg	ctcgcctccg	ccgacccccg	cctgcccgctc	cgggtgggtg	9180
tcgttgggtg	ggggagtgaa	tgggtgctacc	ggtcattccc	tcccgcgtgg	tttgactgtc	9240
tcgcgggtgt	cgcgcttctc	tttcgcgcaa	ccccaccgcc	aaccaccac	cctgctctcc	9300
cggcccggtg	cggtcgacgt	tccggctctc	ccgatgccga	ggggttcggg	atttgtgcgc	9360
gggacggagg	ggagagcggg	taagagaggt	gtcggagagc	tgtcccgggg	cgacgctcgg	9420
gttggctttg	ccgcgtgcgt	gtgctcgcgg	acgggttttg	tcggaccccc	acggggtcgg	9480
tccggccgca	tgcactctcc	cgttccgcgc	gagcgcgcgc	ccggctcacc	cccgtttgt	9540
cctcccgcca	ggctctccgc	cgcgcgcgc	tctcctcct	ctctcgcgt	ctctgtcccg	9600
cctggtcctg	tcccaccccc	gacgctccgc	tcgcgcttcc	ttacctgggt	gatcctgcca	9660
ggtagcatat	gcttgtctca	aagattaagc	catgcatgtc	taagtacgca	cggccgggtac	9720
agtgaactg	cgaatggctc	attaaatcag	ttatggttcc	tttgggtcgt	cgtcctctc	9780
ctacttggat	aactgtggtg	attctagagc	taatacatgc	cgacgggcgc	tgacccccct	9840
tcccgggggg	ggatgcgtgc	atttatcaga	tcaaaaccaa	cccgggtgagc	tccctcccg	9900
ctccggccgg	gggtcgggcg	ccggcggtt	ggtgactcta	gataacctcg	ggccgatcgc	9960
acgccccccc	tggcggcgac	gacccattcg	aacgtctgcc	ctatcaactt	tcgatggtag	10020
tcgcgctgcc	taccatgggt	accacgggtg	acggggaatc	aggggttcgat	tccggagagg	10080
gagcctgaga	aacggctacc	acatccaagg	aaggcgcag	gcgcgcaaat	taccactcc	10140
cgacccgggg	aggtagtgc	gaaaaataac	aatacaggac	tctttcgagg	ccctgtaatt	10200
ggaatgagtc	cacttttaaat	cctttaacga	ggatccattg	gagggcaagt	ctgggtgccag	10260
cagccgcggg	aattccagct	ccaatagcgt	atattaaagt	tgtgcagtt	aaaaagctcg	10320
tagttggatc	ttgggagcgg	gcgggcgggt	cgcgcgagg	cgagtacccg	cccgctcccc	10380
ccccttgctc	ctcgggcgcc	cctcgatgct	cttagctgag	tgtcccgcgg	ggcccgagg	10440
gtttactttg	aaaaaattag	agtgttcaaa	gcaggcccg	gcccctgga	taccgcagct	10500
aggaataatg	gaataggacc	gcggttctat	tttgttgggt	ttcggaactg	aggccatgat	10560
taagaggggac	ggccgggggc	attcgtattg	cgcgcgtaga	ggtgaaattc	ttggaccggc	10620

-7-

gcaagacgga	ccagagcgaa	agcattttgcc	aagaatgttt	tcattaatca	agaacgaaag	10680
tccgaggttc	gaagacgatc	agataccgctc	gtagttccga	ccataaacga	tgccgactgg	10740
cgatgcgcgc	gcgttattcc	catgaccgcgc	cgggcagctt	ccgggaaacc	aaagtctttg	10800
ggttccgggg	ggagtatggt	tgcaaagctg	aaacttaaag	gaattgacgg	aagggcacca	10860
ccaggagtgg	gcctgcggct	taatttgact	caacacggga	aacctcacc	ggcccggaca	10920
cggacaggat	tgacagattg	atagctcttt	ctcgattccg	tgggtgggtg	tgcattggccg	10980
ttcttagttg	gtggagcgat	ttgtctggtt	aattccgata	acgaacgaga	ctctggcatg	11040
ctaactagtt	acgcgacccc	cgagcggctc	gcgtccccc	acttcttaga	gggacaagtg	11100
gcgttcagcc	acccgagatt	gagcaataac	aggtctgtga	tgcccttaga	tgtccggggc	11160
tgcacgcgcg	ctacactgac	tggctcagcg	tgtgcctacc	ctgcgcgggc	aggcgccggg	11220
aaccgcgttg	acccattccg	tgatggggat	cggggattgc	aattattccc	catgaacgag	11280
gaattcccg	taagtgcggg	tcataagctt	cgcttgatta	agtcctgcc	ctttgtacac	11340
accgcccgtc	gctactaccg	attggatggt	ttagtgaggc	cctcggatcg	gccccgcggg	11400
ggtcggccca	cggccctggc	ggagcgtga	gaagacggtc	gaacttgact	atctagagga	11460
agtaaaagtc	gtaacaaggt	ttccgtaggt	gaacctgcgg	aaggatcatt	aaacgggaga	11520
ctgtggagga	gcggcggcgt	ggcccgcctc	ccccgtcttg	tgtgtgtcct	cgccgggagg	11580
cgcgtgcgct	cgggtcccg	tcgcccgcgt	gtggagcgag	gtgtctggag	tgaggtgaga	11640
gaaggggtgg	gtggggctcg	tctgggtccg	tctgggaccg	cctccgattt	ccccccccc	11700
tccctctctc	ctcgtccggc	tctgacctcg	ccacctacc	gcggcggcgg	ctgctcgcgg	11760
gcgtcttgcc	tctttcccg	ccggtctctc	cgtgtctacg	aggggcggta	cgtcgttacg	11820
ggtttttgcc	ccgtccggg	ggcgttcggt	cgtcggggcg	cgcgctttgc	tctcccggca	11880
cccatcccg	cgcggctct	ggcttttcta	cgctggctgg	ggcggttgtc	gcgtgtgggg	11940
ggatgtgagt	gtcgcgtgtg	ggctcggccg	tcccgatgcc	acgcttttct	ggcctcgcgt	12000
gtctctcccg	ctcctgtccc	gggtaccctag	ctgtcgcggt	ccggcgcgga	ggtttaagga	12060
ccccgggggg	gtcgccctgc	cgccccagg	gtcggggggc	ggtggggccc	gtagggaagt	12120
cggtcgcttc	ggcggctctc	cctcagactc	catgacctc	ctcccccg	tgccgcggtt	12180
cccgaggcgg	cggctgtgtg	gggggggtgga	tgtctggagc	ccccctgggc	gccgtggggg	12240
cccgacccgc	gccgcggcgt	tgcccgattt	ccgcgggtcg	gtcctgtcgg	tgccgggtcg	12300
gggttcccg	gtcgttcccg	tgtttttccg	ctcccgaccc	tttttttttc	ctccccccca	12360
cacgtgtctc	gtttcgcttc	tgttgcccg	cctgaggcta	ccccctcggtc	catctgttct	12420
cctctctctc	agggagaggg	agggcggtgg	tcgttggggg	actgtgccgt	cgtcagcacc	12480
cgtgagttcg	ctcacaccgc	aaataccgat	acgactctta	gcgggtggatc	actcggctcg	12540
tgcgtcgatg	aagaacgcag	ctagctgcga	gaattaatgt	gaattgcagg	acacattgat	12600
catcgacact	tcgaacgcac	ttgcggcccc	gggttccctc	cggggctacg	cctgtctgag	12660
cgtcggttga	cgatcaatcg	cgtcacccgc	gtcgggtggg	gctgcgcggc	tgggagtttg	12720
ctcgcagggc	caacccccca	accggggtcg	ggccctccgt	ctcccgaagt	tcagacgtgt	12780
gggcgggtgt	cgggtgtggcg	cgcgcgcccc	cgtcgcggag	cctgggtctcc	cccgcgcac	12840
cgcgctcgcg	gcttcttccc	gctccgccc	tcccgccctc	gcccgtgcac	cccggctcctg	12900
gcctcgcgtc	ggcgcctccc	ggaccgctgc	ctcaccagtc	tttctcgggtc	ccgtgccccg	12960
tgggaaccca	cgcgcgcccc	gtggcgcccg	ggggtggggc	cgtccgcac	tgctctggtc	13020
gaggttggcg	gttgaggggtg	tgcgtgcgcc	gaggtgggtg	tccgtcccc	gcggcccgcg	13080
ggttgtcggg	gtggcggtcg	acgagggccg	gtcggctgcc	tgccgtgggt	gtctgtgtgt	13140
gtttgggtct	tgcgctgggg	gagggcgggg	cgaaccgtcg	cggggttggc	gcggtcgccc	13200
ggcgcccgcg	accctccggc	ttgtgtggag	ggcagagcag	ggcgagaacg	gagagaggtg	13260
gtatcccgcg	tggcggtgcg	agggaggggt	tggcgctccg	cgtccgtccg	tccctccctc	13320
cctcgggtgg	cgccttcgcg	ccgcacgcgg	ccgctagggg	cggctcggggc	ccgtggcccc	13380
cgtggctctt	cttcgtctcc	gcttctcctt	caaccggggc	gtaccgcgtc	cggcgccggc	13440
ccgcgggacg	ccgcggcgct	cgtgcgccc	tgcgagtcac	ccccgggtgt	tgcgagttcg	13500
gggagggaga	gggcctcgct	gaccggttgc	gtcccggtt	ccctgggggg	gaccgggcgt	13560
ctgtgggctg	tgcgtcccg	gggttgcgtg	tgagtgaagt	cctccacccc	cgcgcgctc	13620
ccctcccgcc	ggcctctcgg	ggacccctcg	agacggttcg	ccggctcgtc	ctcccggtgc	13680
gcccgggtgc	gtctctttcc	cgcggccctc	ctcgctctct	tcttcccgcg	gctgggcgcg	13740
tgtccccct	ttctgaccgc	gacctcagat	cagacgtggc	gaccgcgtga	atttaagcat	13800
attagtcacg	ggagaaaaag	aaactaacca	ggattccctc	agtaacggcg	agtgaacagg	13860
gaagagccca	gcgcgaatc	ccgcgcgcgc	gtcgcggcgt	gggaaatgtg	gcgtacggaa	13920
gacccactcc	ccggcgccgc	tctggggggg	cccaagtcct	tctgatcgag	gcccagcccc	13980
tggacgggtg	gaggccggta	gcggccccc	cgcgcggggc	tccgggtctt	ccggagtcgg	14040
gttgcttggt	aatgcagccc	aaagcgggtg	gtaaaactca	tctaaggcta	aataccggca	14100
cgagaccgat	agtcaacaag	taccgtaagg	gaaagttgaa	aagaactttg	aagagagagt	14160
tcaagagggc	gtgaaaccgt	taagaggtaa	acgggtgggg	tccgcgcagt	ccgcccggag	14220
gattcaaccc	ggcggcgcg	gtccggccgt	gcccgggtgt	cccggcggat	ctttcccgct	14280
ccccgttcc	cccgacccct	ccacccgcgc	gtcgttcccc	tcttccctcc	cgcgtccggc	14340
gcctccggcg	gcgggcgcgg	ggggtgggtg	ggtgggtggc	cgcggggcgg	gcccgggggtg	14400
gggtcggcg	cgggcggccc	cgggcggcg	accgggcggc	ggcggggcga	cttccaccgt	14460
ggcgggtgcg	cgcgaccggc	tccgggacgg	ccgggaaggc	ccgggtgggga	aggtgggtcg	14520
ggggggggcg	cgcgtctcag	ggcgcgccga	accacctcac	cccagtggtt	acagccctcc	14580
ggccgcgctt	tgcgcaatc	ccggggccga	ggaagccaga	taccgctcgc	cgcgctctcc	14640

-8-

ctctcccccc	gtccgcctcc	cggggggggcg	tggggggtggg	ggcggggccg	cccccccac	14700
ggcgccgacg	ctctcccacc	ccccctcgctc	gcctctctctc	gggcccgggtg	ggggggcgggg	14760
cggactgtcc	ccagtgccgcc	ccggggcgctc	tcgcgcgcgtc	gggtcccggg	gggaccgtcg	14820
gtcacgcgtc	tcccgcacgaa	gccgagcgca	cgggggtcggc	ggcgatgtcg	gctacccacc	14880
cgacccgtct	tgaacacagg	accaaggagt	ctaaccgcgtg	cgcgagtcag	gggctcgtcc	14940
gaaagccgcc	gtggcgcaat	gaaggtgaag	ggccccgccc	ggggggcccg	ggtgggatacc	15000
cgaggcctct	ccagtcgcgc	gagggcgcac	caccggcccg	tctcgcccgc	cgcgcgggg	15060
aggtggagca	cgagcgtacg	cgttaggacc	cgaaagatgg	tgaactatgc	ttgggacggg	15120
cgaagccaga	ggaaactctg	gtggaggtcc	gtagcgggtcc	tgacgtgcaa	atcggtcgtc	15180
cgacctgggt	ataggggcga	aagactaatc	gaaccatcta	gtagctggtt	ccctccgaag	15240
tttccctcag	gatagttggc	gctctcgctc	ccgacgtacg	cagttttatc	cggtaaagcg	15300
aatgattaga	ggtcttgggg	ccgaaacgat	ctcaacctat	tctcaaaactt	taaatgggta	15360
agaagcccg	ctcgctggcg	tggagccggg	cgtggaatgc	gagtgccctag	tggggccactt	15420
ttggtaagca	gaactggcgc	tgcgggatga	accgaacgcc	gggttaaggc	gcccgatgcc	15480
gacgctcatc	agaccccaga	aaaggtgttg	gttgatatag	acagcaggac	ggtggccatg	15540
gaagtccgaa	tccgctaagg	agtgtgtaac	aactcacctg	ccgaatcaac	tagccctgaa	15600
aatggatggc	gctggagcgt	cgggcccata	cccggccgtc	gccgcagtcg	gaacgggaacg	15660
ggacgggagc	ggccgcgggt	gcgcgtctct	cgggggtcggg	ggtgctgggc	gggggcccgt	15720
cccccgccct	ccctccgcgc	gccgggttcg	cccccgcgcc	gtcgggcccc	gcccagccct	15780
cgcgcgcacg	agtaggaggg	ccgctgcggt	gagccttgaa	gcctaggggc	cggggccggg	15840
tggagccgcg	gcaggtgcag	atcttgggtg	tagtagcaaa	tattcaaacg	agaactttga	15900
agggcgaagt	ggagaaagggt	tccatgtgaa	cagcagttga	acatgggtca	gtcgggtctg	15960
agagatgggc	gagtgccgtt	ccgaaggagc	gggcgatggc	ctccgttgcc	ctcgccgat	16020
cgaaggggag	tcgggttcag	atccccgaat	ccggagtggc	ggagatgggc	gccgcgaggc	16080
cagtgcggta	acgcgaccga	tcccggagaa	gccggcgggg	ggcctcgggg	agagttctct	16140
tttctttgtg	aagggcaggg	cgccctggaa	tgggttcgcc	ccgagagagg	ggccgatgcc	16200
ttggaaagcg	tcgcggttcc	ggcggcgtcc	ggtgagctct	cgctggccct	tgaaaatccg	16260
ggggagaggg	tgtaaatctc	gcgcggggcc	gtacccatat	ccgcagcagg	tctccaagg	16320
gaacagcctc	tggcatgttg	gaacaatgta	ggtaagggaa	gtcggcaagc	cggatccgta	16380
acttcgggat	aaggattggc	tctaagggct	gggtcgggtc	ggctggggcg	cgaagcgggg	16440
ctgggcgcgc	gccgcggctg	gacgagcgcg	cgccgccctc	tcccacgtcc	ggggagaccc	16500
cccgctcttt	ccgcccgggc	ccgcctccc	ctcttcccgc	cggggccccg	tcgtcccccg	16560
cgtcgtcgcc	acctctcttc	ccccctcctt	cttcccgtcg	gggggcgggt	cgggggtcgg	16620
cgcgcggcgc	gggctccggg	gcggcgggtc	caaccccgcg	ggggttccgg	agcgggagga	16680
accagcggtc	cccggtgggg	cggggggccc	ggacactcgg	ggggccggcg	gcggcgccga	16740
ctctggacgc	gagccgggcc	cttcccgtgg	atgcctcag	ctgcggcggg	cgtcgccgcc	16800
gctcccgggg	agcccggcgg	gtgccggcgc	gggtcccctc	cccgcggggc	ctcgctccac	16860
ccccccatcg	cctctcccga	ggtgcgtggc	ggggcggggc	gggcgtgtcc	cgcgcgtgtg	16920
gggggaacct	ccgcgtcggt	gttccccccg	cgggtccgcc	ccccggggcg	cggttttccg	16980
cgcggcgccc	ccgcctcggc	cgggcctag	cagccacttc	agaactgggt	cggagacggg	17040
gaatccgact	gtttaattaa	aacaaagcat	cgcgaaggcc	cgcggcgggg	gttgacgcga	17100
tgtgatttct	gcccagtgct	ctgaatgtca	aagtgaagaa	attcaatgaa	gcgcgggtaa	17160
acggcgggag	taactatgac	tctcttaagg	tagccaaatg	cctcgtcatc	taattagtga	17220
cgcgcataaa	tggatgaacg	agattcccac	tgtcccctacc	tactatccag	cgaaacacca	17280
gccaaaggaa	cgggcttggc	ggaatcagcg	gggaagaaga	accctgttga	gcttgactct	17340
agtctggcac	ggtgaagaga	catgagaggt	gtagaataag	tgggaggccc	ccggcgcccc	17400
gccccgtcct	cgcgtcgggg	tgggggcacg	ccggcctcgc	gggcccggcg	tgaaatacca	17460
ctactctcat	cgttttttca	ctgacccggg	gaggcggggg	ggcgagcccc	gaggggctct	17520
cgttctgggc	gccaaagcgtc	cgtcccgcgc	gtgcggggcg	gcgcgacccg	ctccggggac	17580
agtgccagggt	ggggagtttg	actggggcgg	tacacctgtc	aaacggtaac	gcaggtgtcc	17640
taaggcgagc	tcagggagga	cagaaacctc	ccgtggagca	gaagggcaaa	agctcgcttg	17700
atcttgattt	tcagtacgaa	tacagaccgt	gaaagcgggg	cctcacgata	cttctgacct	17760
tttgggtttt	aagcaggagg	tgtcagaaaa	gttaccacag	ggataactgg	cttgtggcgg	17820
ccaagcgttc	atagcgacgt	cgttttttga	tccttcgatg	tcggctcttc	ctatcattgt	17880
gaagcagaat	tcaccaagcg	ttggattgtt	caccacttaa	tagggaaactg	gagctggggt	17940
tagaccgtcg	tgagacaggt	tagttttacc	ctactgatga	tgtgttgttg	ccatggtaat	18000
cctgctcagt	acgagaggaa	ccgcaggttc	agacatttgg	tgtatgtgct	tggctgagga	18060
gccaatgggg	cgaagctacc	atctgtggga	ttatgactga	acgcctctaa	gtcagaatcc	18120
gcccgaagcg	aacgatacgg	cagcgccgaa	ggagcctcgg	ttggcccccg	atagccgggt	18180
ccccgtccgt	cccgctcggc	gggggtcccc	cgtgcccccg	cggcgccgcg	gggtctcccc	18240
ccgcggggcg	tcgggacccg	ggtccgggtg	ggagagccgt	tcgtcttggg	aaacgggggtg	18300
cggccggaaa	ggggggccgc	ctctcgcccc	tcacgttgaa	cgcacgttcg	tgtggaaacct	18360
ggcgctaaac	cattcgtaga	cgacctgctt	ctgggtcggg	gtttcgtagc	tagcagagca	18420
gctccctcgc	tgcgatatat	tgaaagtcag	ccctcgacac	aagggtttgt	ctctcggggc	18480
tttcccgctc	cacgcccgtc	cgtctgcacg	cgacctgttc	gccgccccgg	cgtcacgggg	18540
gcggtcgcct	cggcccccg	gcggttgccc	gaacgaccgt	gtgggtgggtg	gggggggggat	18600
cgtcttctcc	tccgtctccc	gaggacgggt	cgtttctctt	tccccctccg	tcgctctcct	18660

-9-

tggtgtgtgtgt	agcctcgtgc	cgtcgcgacc	gcggcctgcc	gtcgcctgcc	gccgcagccc	18720
cttgccctcc	ggccttgggc	aagccggagg	gcggaggagg	gggatcggcg	gcggcggcga	18780
ccgcggcgcg	gtgacgcacg	gtgggagccc	catcctcggc	gcgtccgtcg	gggacggccg	18840
gttgaggagg	cgggaggggg	ttttcccggt	aacgcgcgct	tcggcgccag	gcctctggcg	18900
gccggggggg	cgctctctcc	gcccgagcat	ccccactccc	gccccctctc	ttcgcgcgcc	18960
gcggcgggga	cgtgcgtacg	aggggaggat	gtcgcgggtg	ggaggcggag	aggggtccggc	19020
gcggcgccctc	ttccattttt	ttccccccaa	cttcggaggt	cgaccagtac	tcggggcgac	19080
actttgtttt	ttttttttcc	cccgatgctg	gaggctcgacc	agatgtccga	aagtgtcccc	19140
cccccccccc	ccccccggcg	cggagcggcg	ggggcactct	ggactctttt	tttttttttt	19200
tttttttttt	ttaaattcct	gggaaccttta	ggtcgaccag	ttgtccgtct	tttactcctt	19260
catataggtc	gaccagtact	ccgggtggta	ctttgtcttt	ttctgaaaa	cccagaggtc	19320
gaccagatat	ccgaaagtcc	gttgtccgtc	tttactcttc	cccacagcga	ttctcttttt	19380
tttttttttt	tttggtgtgc	ctctttttga	cttatataca	tgtaaatagt	gtgtacgttt	19440
atatacttat	aggaggagg	cgaccagtac	tcggggcgac	actttgtttt	tttttttttt	19500
tcacccgatg	atggagggtcg	accagatgtc	cgaaagtgtc	ccgtccccc	cctccccccc	19560
ccgcgacgcg	gcgggctcac	tctggactct	tttttttttt	tttttttttt	tttaaaatttc	19620
tggaacctta	aggtcgacca	gttgtccgtc	tttactctat	tcataatagg	cgaccgggtg	19680
tactttgtct	ttttctgaaa	atcgcagagg	tcgaccagat	gtcagaaagt	ctgggtggctg	19740
ataaattatc	tgatctagat	ttgtttttct	gtttttcagt	tttgtgttgt	tttgtgttgt	19800
tttgtgttgt	tttgttttgt	tttgttttgt	tttgttttgt	tttgttttgt	tttgttttgt	19860
tttgtgttgt	gttgtgttgt	gttgtgttgt	gttgggttgg	gttgggttgg	gttgggttgg	19920
gttgggttgg	gttgggttgg	gttgggttgg	tttgtgttgt	tttgtgttgt	tttgtgttgt	19980
ttgtttgtctg	ttgtttgtgt	tttgcgggtg	cgaacagtgt	tccttaaccg	agtttttttg	20040
tacacaaaca	tgcaactttt	ttaaaataaa	tttttaaaat	aaatgcgaaa	atcgaccaat	20100
tatccctttc	cttctctctc	ttttttaaaa	attttctttg	tgtgtgtgtg	tgtgtgtgtg	20160
tgtgtgtgtg	tgcgtgtgtg	tgtgtgtgtg	cgtgcagcgt	gcgcgcgcct	gtttttataaa	20220
tactatatag	aataaggtcg	cgggtgtgtg	tagcttcccg	gactccagag	gcagaggcag	20280
gcagacttct	gagttcgagg	ccagcctggg	ctacagagga	accctgtctc	gaaaaatgaa	20340
aataaataca	tacatacata	catatacata	tacatacata	catatacata	tacatatgag	20400
gttgaccagt	tgtcaatcct	ttagaatttt	gtttttaatt	aatgtgatag	agagatagat	20460
aatagataga	tggatagagt	gatacaataa	taggtttttt	tttcagttaa	tatgagggtg	20520
attaaccact	tttccctttt	taggtttttt	tttttttccc	ctgtccatgt	gggtgtgtgg	20580
atttgaactc	aggaccctgg	cagggtcaact	ggaaaacgtg	ttttctatat	atataaatag	20640
tggtctgtct	gctgttttgt	tgtttgtctt	cttgcctgtc	tgcttgcctg	cttgcctgtc	20700
tgcctttttt	tttcttctga	gacagtattt	ctctgtgtaa	cctgggtgcc	tgaacctcac	20760
tctgtatgac	agcctggcct	caatcgaaat	cagaaatcct	cctgcctctt	gtctacctcc	20820
caatttttga	gtaaagggtg	gctacaccac	tgcttggcat	tattatcatt	atcattatta	20880
attttattat	tagacagaac	gaaatcaact	agttggctct	gtttcgttaa	ttcatttgaa	20940
attagtgtga	ccaattagtt	ggctgggttt	ggagggtttc	tttgtttccg	atttgggtgt	21000
ttgtggggct	ggggatcagg	tatctcaacg	gaatgcatga	aggtaagggt	gagatggctc	21060
gatttttgtga	aagattactt	ttcttagtct	gaggaaaaaa	taaaaataa	ttggggtacg	21120
tttcattgtc	tcatttctat	ttctctttct	ttctttcttt	ctttcagata	aggagggtcg	21180
ccagttcctc	ctgccttctg	gaagatgtag	gcattgcatt	gggaaaagca	ttgtttgaga	21240
gatgtgctag	tgaaccagag	agtttggatg	tcaagccgta	taatgtttat	tacaatatag	21300
aaaagtctta	acaaagtgat	ctttaacttt	tttttttttt	tttctccttc	tacttctact	21360
tgttctcact	ctgccaccac	cgcgctttgt	acattgaatg	tgagctttgt	tttgcctaac	21420
agacatata	tttttctttt	ggttttgtct	gacatgggtt	ccctttctat	ccgtgcaggg	21480
ttcccagacg	gccttttgag	aataaaatgg	gaggccagaa	ccaaagtctt	ttgaataaag	21540
caccacaact	ctaactgttt	tggtgttttt	ccttcccaag	gcacagatct	ttcccagcat	21600
ggaaaagcat	gtagcagttg	taggacacac	tagacgagag	caccagatct	cattgtgggt	21660
ggttgtgaac	caccaccat	gtggttgctt	gggatttgaa	ctcaggatct	tcagaagacg	21720
agtcagggtc	ctaaaccgat	gagccatctc	tccagccctc	ctacattcct	tcttaaggca	21780
tgaatgatcc	cagcatggga	agacagtctg	ccctctttgt	ggtatatcac	catatactca	21840
ataaaaatac	gaaatgaatg	aagtctccac	gtatttattt	cttcgagcta	tctaaattct	21900
ctcacagcac	ctccccctcc	cccacactgc	ctttctccct	atgtttgggt	gggggtgggg	21960
gaggggtggg	gtggggggcag	ggatctgcat	gtcttcttgc	aggtctgtga	actatttgcg	22020
atggcctggt	tctctgaact	gttgagcctt	gtctatccag	aggtgactg	gctagttttc	22080
tacctgaagt	ccctgagtga	tgatttccct	gtgaattc			22118

<210> 19

<211> 175

<212> DNA

<213> Mus musculus

<400> 19

ctccccgcgcg	gcccccggtg	tcgccgttcc	cgtggcgcg	acaatgcggg	tgtgcgtcca	60
cgtgtgcgtg	tccgtgcagt	gcggttgtgg	agtgccctgc	tctcctctcc	ctccccggca	120

-10-

gcgttcccac ggttggggac caccggtgac ctgcacctct tggggcctgg atccg 175

<210> 20
 <211> 755
 <212> DNA
 <213> Mus musculus

<400> 20
 ggtctggttg gaattgttga cctcgctctc ggggtcgggc tttggggaac ggcggggctg 60
 gtcgtgcccg gcgcgcggac tgtgtcgggg ccacttccc gctcagagggt ggcggtggcg 120
 gcggcggttg tagtctccc tgttgctct tcccgggctc ttgggggggg tgccgtcgtt 180
 ttcggggccg gcgttgcttg gcttacgcag gcttggtttg ggactgcctc aggagtcgtg 240
 ggcggtgtga ttcgcgcgg ttttgctctg cgtctgcctg ctttgctctg ggtttgcttg 300
 gttcgtgtct cgggagcggg ggtttttttt tttttcgggt cccggggaga ggggtttttc 360
 cgggggacgt tcccgtcgcc cctgcgcgc ggtgggtttt cgtttcgggc tgtgttcgtt 420
 tccccttccc cgtttcgccg tcggttctcc ccggtcggtc ggccctctcc ccggtcggtc 480
 gcccgggccg gctgcgggac cccccttct gggggggatg cccgggcacg caccgtccg 540
 ggcgggccact gtggtccggg agctgctcgg caggcgggtg agccagttag aggggcgtca 600
 tgcccccgcg ggctcccgtg gccgacgcgg cgtgttcttt gggggggcct gtgcgtgcgg 660
 gaaggctgcg cacgttgtcg gtccctgcga gggaaagagg cttttttttt ttaggggggtc 720
 gtccttcgtc gtcccgtcgg cgggtgatcc ggcct 755

<210> 21
 <211> 463
 <212> DNA
 <213> Mus musculus

<400> 21
 ggccgaggtg cgtctcgagg ttggggctcg tccggccccc togtcctccg ggaaggcgtt 60
 tagcgggtac cgtcgccgcg ccgaggtggg cgcacgtcgg tgagataacc ccgagcgtgt 120
 ttctgggttg ttggcgccgg ggtcccggtc gatgtcttcc cctccccctc tcccgcaggc 180
 caggtcagcc tccgcctgtg ggcttcgtcg gccgtctccc cccccctcac gtccctcgcg 240
 agcgagcccg tccgttcgac cttccttccg ccttcccccc atctttccgc gctccgttgg 300
 ccccggggtt ttacagggcg cccccacgct cctccgcctc tccgcccgtg gtttggacgc 360
 ctggttccgg tctccccgcg aaaccccggt tgggttggtc tccggccccg gcttgctctt 420
 cgggtctccc aacccccggc cgggaagggt cgggggttcc ggg 463

<210> 22
 <211> 378
 <212> DNA
 <213> Mus musculus

<400> 22
 ggattcttca ggattgaaac ccaaaccggg tcagtttctt ttccggctcc ggcggggggg 60
 ggcggccccc ggcggttttg tgagttagat aacctcgggc cgatcgacg ccccccgtag 120
 cggcgacgac ccattcgaac gtctgccta tcaactttcg atggtagtcg atgtgcctac 180
 catggtgacc acgggtgacg gggaatcagg gttcgattcc ggagagggag cctgagaaac 240
 ggctaccaca tccaagggaag gcagcaggcg cgcaaattac ccactcccga cccggggagg 300
 tagtgacgaa aaataacaat acaggactct ttcgaggccc tgtaattgga atgagtccac 360
 tttaaatcct ttaagcag 378

<210> 23
 <211> 378
 <212> DNA
 <213> Mus musculus

<400> 23
 gatccattgg agggcaagtc tgggtgccagc agccgcggta attccagctc caatagcgtg 60
 tattaagatt gctgcagtta aaaagctcgt agttggatct tgggagcggg cgggcgggtc 120
 gccgcgaggg gagtcaccgc ccgtccccgc cccttgctc tcggcgcccc ctcatgctc 180
 ttagctgagt tgtcccgcg ggccogaagc gtttactttg aaaaaattag agttgtttca 240
 aagcaggccc gagccgcctg gataccgcca gctaggaaat aatggaatag gaccgcgggt 300
 cctattttgt ttggttttgc gaactgagcc catgattaag ggaaacggcc gggggcattc 360
 ccttattgcg ccccccta 378

<210> 24
 <211> 719

-11-

<212> DNA

<213> Mus musculus

<400> 24

ggatcttttcc	cgctccccgt	tcttccccggc	ccctccacccc	gcgcgtctcc	cccccttcttt	60
tcccctctcc	ggagggggggg	gaggtggggg	cgcgtgggcg	gggtcggggg	tgggggtcggc	120
gggggaccgc	ccccggcccg	caaaaggccg	ccgcggggcg	cacttcaacc	gtacgggtgc	180
gccgcgaccg	gctacgagac	ggctgggaag	gcccgcgagg	gaatgtggct	cgggggggggc	240
ggcgcgtctc	agggcgcgcc	gaaccacctc	accccgagtg	ttacagccct	ccggcccgcc	300
tttcgcggaa	tccccggggc	gaggggaagc	ccgatacccc	tcgcgcgct	tttccccctcc	360
ccccgtccgc	ctcccggggc	ggcgtggggg	tgggggcccg	gccgcccctc	ccacgcccg	420
ggtttctctc	tctcccggtc	tcggccgggt	tggggggggg	agcccgggtg	ggggcgggggc	480
ggactgtcct	cagtgcgccc	cgggcgctcg	cgcgcgctcg	ggcccggggg	gttctctcgg	540
tcacgcccgc	cccgcgaag	ccgagcgcac	ggggtcggcg	gcgatgtcgg	ctacccaacc	600
gaccggtctt	gaaacacgga	ccaaggagtc	taacgcgtgc	gcgagtcagg	ggctcgcacg	660
aaagccgcgc	tggcgcaatg	aaggtgaagg	gcccgtccg	ggggcccag	gtgggatcc	719

<210> 25

<211> 685

<212> DNA

<213> Mus musculus

<400> 25

cgaggcctct	ccagtccgcc	gagggcgcac	caccggcccc	tctcgccgc	cgcgtcgggg	60
aggtggagca	cgagcgtacg	cgttaggacc	cgaaagatgg	tgaactatgc	ctgggcaggg	120
cgaagccaga	ggaaactctg	gtggaggtcc	gtagcgggtc	tgacgtgcaa	atcggtcgtc	180
cgacctgggt	ataggggcga	aagactaatc	gaaccatcta	gtagctgggt	ccctccgaag	240
tttccctcag	gatagctggc	gctctcgcaa	ccttcggaag	cagttttatc	cggggtaaagg	300
cggaatggat	taggaggtct	tggggccgga	aacgatctca	aactatttct	caaactttaa	360
atgggtaaag	aagcccggct	cgctggcgctg	gagccggggc	tggaaatgca	gtgcctagt	420
ggccactttt	ggtaagcaga	actggcgctg	cgggatgaac	cgaacgcggg	gttaaggcgc	480
ccgatgccga	cgctcatcag	accccgaaaa	aggtgttggt	tgatatagac	agcaggacgg	540
tggccatgga	agtccgaatc	cgctaaggag	tgtgtaacaa	ctcacctgcc	gaatcaacta	600
gccctgaaaa	tggatggcgc	tggagcgctc	ggcccatacc	cggccgtcgc	cggcagtcgg	660
aacgggacgc	gacgggagcg	gccgc				685

<210> 26

<211> 5162

<212> DNA

<213> Artificial Sequence

<220>

<223> Chimeric bacterial plasmid

<400> 26

gacggatcgg	gagatctccc	gatcccctat	ggtcgactct	cagtacaatc	tgctctgatg	60
ccgcatagtt	aagccagtat	ctgctccctg	cttgtgtgtt	ggagggtcgt	gagtagtgcg	120
cgagcaaaat	ttaagctaca	acaaggcaag	gcttgaccga	caattgcatg	aagaatctgc	180
ttagggttag	gcgttttgcg	ctgcttcgcg	atgtacgggc	cagatatacg	cgttgacatt	240
gattattgac	tagttattaa	tagtaatcaa	ttacggggtc	attagttcat	agcccatata	300
tggagttccg	cgttacataa	cttacggtaa	atggccccgc	tggctgaccg	cccaacgacc	360
cccgcccat	gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	420
attgacgtca	atgggtggac	tatttacggg	aaactgcccc	cttggcagta	catcaagtgt	480
atcatatgcc	aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	540
atgccagta	catgacctta	tgggactttc	ctacttgcca	gtacatctac	gtattagtca	600
tcgctattac	catggtgatg	cggttttgcc	agtacatcaa	tgggcgtgga	tagcggtttg	660
actcacgggg	atttccaagt	ctccaccccc	ttgacgtcaa	tgggagtttg	ttttggcacc	720
aaaatcaacg	ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	780
gtaggcgtgt	acgggtgggag	gtctatataa	gcagagctct	ctggctaact	agagaaccca	840
ctgcttactg	gcttatcgaa	attaatacga	ctcactatag	ggagacccaa	gcttggtacc	900
gagctcggag	cgatatctgc	ggccgcgctc	acggaattca	gtggatccac	tagtaacggc	960
cgccagtggt	ctggaattaa	ttcgctgtct	gcgagggcca	gctggtgggg	tgagtaactcc	1020
ctctcaaaag	cgggcatgac	ttctgcgcta	agattgtcag	tttccaaaaa	cgaggaggat	1080
ttgatattca	cctggcccgc	ggtgatgcct	ttgagggtgg	ccgcgtccat	ctggtcagaa	1140
aagacaatct	ttttgttgtc	aagcttgagg	tgtggcaggc	ttgagatctg	gccatacact	1200
tgagtgacaa	tgacatccac	tttgcctttc	tctccacagg	tgtccactcc	caggtccaac	1260
tgcaggtcga	gcattgcatt	agggcgccca	attccgcccc	tctccctccc	ccccccctaa	1320

-12-

cgttactggc	cgaagccgct	tgggaataagg	ccgggtgtgcy	tttgtctata	tgtgattttc	1380
caccatattg	ccgtcttttg	gcaatgtgag	ggcccggaaa	cctggccctg	tcttcttgac	1440
gagcattctg	aggggtcttt	cccctctcgc	caaaggaatg	caaggctctg	tgaatgtcgt	1500
gaaggaagca	gttcctctgg	aagcttcttg	aagacaaaac	acgtctgtag	cgaccctttg	1560
caggcagcgg	aacccccccac	ctggcgacag	gtgcctctgc	ggccaaaagc	cacgtgtata	1620
agatacacct	gcaaaggcgg	cacaacccca	gtgccacgtt	gtgagttgga	tagttgtgga	1680
aagagtcaaa	tggctctcct	caagcgtatt	caacaagggg	ctgaaggatg	cccagaagg	1740
accccatgtg	atgggatctg	atctggggcc	tccgtgcaca	tgctttacat	gtgttttagtc	1800
gagggtaaaa	aaacgtctag	gccccccgaa	ccacggggac	gtggttttcc	tttgaaaaac	1860
acgatgataa	gcttgccaca	accggggatc	caccggtcgc	caccatgggtg	agcaaggggc	1920
aggagctggt	caccgggggtg	gtgcccaccc	tggtcgagct	ggacggcgac	gtaaacggcc	1980
acaagttcag	cgtgtccggc	gagggcgagg	gcgatgccac	ctacggcaag	ctgaccctga	2040
agttcatctg	caccaccggc	aagctgcccc	tgccctggcc	caccctcggtg	accaccctga	2100
cctacggcgt	gcagtgtctc	agccgtacc	ccgaccacat	gaagcagcac	gacttcttca	2160
agtccgccat	gcccgaaggc	tacgtccagg	agcgcaccat	cttcttcaag	gacgacggca	2220
actacaagac	ccgcgccgag	gtgaagtctg	agggcgacac	cctggtgaac	cgcatcgagc	2280
tgaagggtcat	cgacttcaag	gaggacggca	acatcctggg	gcacaagctg	gagtacaact	2340
acaacagcca	caacgtctat	atcatggccg	acaagcagaa	gaacggcatc	aagggtgaact	2400
tcaagatccg	ccacaacatc	gaggacggca	gcgtgcagct	cgccgaccac	taccagcaga	2460
acacccccat	cggcgacggc	cccgtgctgc	tgcccgcaca	ccactacctg	agcaccaggt	2520
ccgcccctgag	caaaagcccc	aacgagaagc	gcgatcacat	ggtcctgctg	gagttcgtga	2580
ccgcgcggcg	gatcactctc	ggcatggacg	agctgtacaa	gtaaagcggc	cctagagctc	2640
gctgatcagc	ctcgactgtg	ccctctagttg	ccagccatct	gttgtttgcc	cctcccccg	2700
gccttccctg	accctggaag	gtgcccactcc	cactgtcctt	tcctaataaa	atgaggaaat	2760
tgcacgcat	tgtctgagta	ggtgtcattc	tattctgggg	ggtggggtgg	ggcaggacag	2820
caagggggag	gattgggaag	acaatagcag	gcctgctggg	gatgcgggtg	gctctatggc	2880
ttctgaggcg	gaaagaacca	gctggggctc	gagtgcatct	tagttgtggg	ttgtccaaac	2940
tcatcaatgt	atcttatcat	gtctgtatac	cgtcgacctc	tagctagagc	ttggcgtaat	3000
catggtcata	gctgtttcct	gtgtgaaatt	gttatccgct	cacaattcca	cacaacatac	3060
gagccggaag	cataaagtgt	aaagcctggg	gtgcctaatt	agtgagctaa	ctcacattaa	3120
ttgcgttgcg	ctcactgccc	gctttccagt	cgggaaacct	gtcgtgccag	ctgcattaat	3180
gaatcggccca	acgcgcgggg	agaggcggtt	tgcgtatttg	gcgctcttcc	gcttccctgc	3240
tactgactc	gctgcgctcg	gtcgttcggc	tgccgagcag	ggtatcagct	cactcaagg	3300
cggtaatcag	gttatccaca	gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaag	3360
ggcagcaaaa	ggccaggaac	cgtaaaaagg	ccgcgttgct	ggcgtttttc	cataggctcc	3420
gccccctga	cgagcatcac	aaaaatcgac	gctcaagtca	gagggtggcg	aaaccgcagc	3480
gactataaag	ataccaggcg	tttccccctg	gaagctccct	cgtgcgctct	cctgttccga	3540
ccctgccgct	taccggatag	ctgtccgctc	ttctcccttc	gggaagcgtg	gcgctttctc	3600
aatgctcacg	ctgtaggtat	ctcagttcgg	tgtaggtcgt	tcgctccaag	ctgggctgtg	3660
tgcacgaacc	cccgcgttcag	cccgaccctc	gcgccttatc	cggtaactat	cgctcttagt	3720
ccaaccgggt	aagacacgac	ttatcgccac	tgccagcagc	cactggtaac	aggattagca	3780
gagcgaggta	tgtaggcggt	gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	3840
ctagaaggac	agtatttggt	atctgcgctc	tgctgaagcc	agttaccttc	ggaaaaagag	3900
ttggtagctc	ttgatccggc	aaacaaaacca	ccgctggtag	cgggtggttt	tttgtttgca	3960
agcagagat	tacgcccaga	aaaaaaggat	ctcaagaaga	tcctttgatc	ttttctacgg	4020
ggtctgacgc	tcagtggaa	gaaaaactcac	gttaagggat	tttgggtcatg	agattatcaa	4080
aaaggatctt	cacctagatc	cttttaaat	aaaaatgaag	ttttaaatca	atctaaagta	4140
tatatgagta	aacttggtct	gacagttacc	aatgcttaat	cagtgaaggca	cctatctcag	4200
cgatctgtct	atttcgttca	tccatagttg	cctgactccc	cgctcgttag	ataactacga	4260
tacgggaggg	cttaccatct	ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	4320
cggtccaga	tttatcagca	ataaaaccagc	cagccggaag	ggccgagcgc	agaagtggtc	4380
ctgcaacttt	atccgcctcc	atccagttca	tttaattgtg	ccgggaagct	agagtaagta	4440
gttcgccagt	taatagtttg	cgcaacggtg	ttgccattgc	tacaggcatc	gtggtgtcac	4500
gctcgtcgtt	tggtatggct	tcattcagct	ccggttccca	acgatcaagg	cgagttacat	4560
gatcccccat	gttgtgcaaa	aaagcgggtta	gctccttcgg	tcctccgac	gttgctcagaa	4620
gtaagttggc	cgcagtgtta	tactcatggg	ttatggcagc	actgcataat	tctcttactg	4680
tcatgccatc	cgtaagatgc	ttttctgtga	ctgggtgagta	ctcaaccaag	tcattctgag	4740
aatagtgtat	gcggcgacccg	agttgtctct	gcccggcgtc	aatacgggat	aataccgcgc	4800
cacatagcag	aaactttaaaa	gtgctcatca	ttggaaaacg	ttcttcgggg	cgaaaaactc	4860
caaggatctt	accgctgttg	agatccagtt	cgatgtaac	cactcgtgca	cccaactgat	4920
cttcagcatc	ttttactttc	accagcgttt	ctgggtgagc	aaaaacagga	aggcaaaatg	4980
ccgcaaaaaa	gggaataagg	gcgacacgga	aatgttgaat	actcatactc	ttcctttttc	5040
aatattattg	aagcatttat	caggggtatt	gtctcatgag	cggatacata	tttgaatgta	5100
tttagaaaaa	taaacaaata	ggggttccgc	gcacatttcc	ccgaaaagtg	ccacctgacg	5160
tc						5162

-13-

<211> 5627

<212> DNA

<213> Artificial Sequence

<220>

<223> pMG plasmid from InvivoGen; IRES sequence modified
EMCV nucleotides 2736-3308

<400> 27

caccggcgaa	ggaggcctag	atctatcgat	tgtacagcta	gctcgacatg	ataagataca	60
ttgatgagtt	tggacaaacc	acaactagaa	tgcagtga	aaaatgcttt	atgtgtgaaa	120
tttgtgatgc	tattgcttta	tttgtgaaat	tttgtatgct	attgctttat	ttgtaaccat	180
tataagctgc	aataaacaag	ttaacaacaa	caattgcatt	cattttatgt	ttcagggttca	240
gggggaggtg	tgggaggttt	tttaagcaa	gtaaaacctc	tacaaatgtg	gtagatccat	300
ttaaatgtta	attaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	360
ggcgcggttg	ctggcggttt	tccataggct	ccgccccctt	gacgagcatc	acaaaaatcg	420
acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	agataccagg	cgtttccccc	480
tggagctcc	ctcgtgcgct	ctcctgttcc	gaccctgccg	cttaccggat	acctgtccgc	540
ctttctccct	tcgggaagcg	tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	600
ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	cccccggtc	agcccgaccg	660
ctgcgcctta	tccggtaact	atcgtcttga	gtccaacccg	gtaagacacg	acttatcgcc	720
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	780
gttcttgaag	tgggtggccta	actacggcta	cactagaaga	acagtatttg	gtatctgcgc	840
tctgctgaag	ccagttacct	tcggaaaaag	agttggtagc	tcttgatccg	gcaaacaaac	900
caccgctggc	agcgggtggc	tttttggttg	caagcagcag	attacgcgca	gaaaaaaagg	960
atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	gctcagtggg	acgaaaactc	1020
acgttaaggg	attttggtca	tggctaagtta	attaagctgc	aataaacaat	cattattttc	1080
attggatctg	tgtgttggtt	ttttgtgtgg	gcttggggga	gggggagggc	agaatgactc	1140
caagagctac	aggaaggcag	gtcagagacc	ccactggaca	aacagtggct	ggactctgca	1200
ccataacaca	caatcaacag	gggagtgcgc	tggatcgagc	tagagtccgt	tacataactt	1260
acggtaaatg	gcccgctggg	ctgaccgccc	aacgaccccc	gcccattgac	gtcaataatg	1320
acgtatgttc	ccatagtaac	gccaataggg	actttccatt	gacgtcaatg	ggtggagtat	1380
ttacggtaaa	ctgcccactt	ggcagtagat	caagtgtatc	atatgccaa	tacgccccct	1440
attgagctca	atgacggtaa	atggcccgcc	tggcattatg	cccagtagat	gaccttatgg	1500
gactttccta	cttggcgagta	catctacgta	ttagtcatcg	ctattaccat	ggtgatgcgg	1560
ttttggcagt	acatcaatgg	gcgtggatag	cggtttgact	cacggggatt	tccaagtctc	1620
caccccatgt	acgtcaatgg	gagtttgttt	tggcaccaaa	atcaacggga	ctttccaaaa	1680
tgtcgttaaca	actccgcccc	attgacgcaa	atgggcggta	ggcgtgtacg	gtgggaggtc	1740
tatataagca	gagctcggtt	agtgaaccgt	cagatcgcc	ggagacgcca	tccacgctgt	1800
tttgacctcc	atagaagaca	ccgggaccga	tccagcctcc	gcggccggga	acggtgcatt	1860
ggaacgcgga	ttccccgtgc	caagagtgc	gtaagtaccg	cctatagagt	ctataggccc	1920
acccccctgg	cttccttatgc	atgctatact	gtttttggct	tgggggtctat	acacccccgc	1980
ttcctcatgt	tcataggtgat	ggtatagctt	agcctatagg	tgtgggttat	tgaccattat	2040
tgacctatcc	ctattgggtg	acgatacttt	ccattactaa	tccataacat	ggctctttgc	2100
cacaactctc	tttattggct	atatgccaat	acactgtcct	tcagagactg	acacggactc	2160
tgtattttta	caggatgggg	tctcatttat	tatttacaaa	ttcacatata	caacaccacc	2220
gtccccagtg	cccgcagttt	ttattaaaca	taacgtggga	tctccacgcg	aatctcgggt	2280
acgtgttccg	gacatgggct	cttctccggt	agcggcggag	cttctacatc	cgagccctgc	2340
tcccagctct	ccagcgactc	atggtcgctc	ggcagctcct	tgctcctaac	agtggaggcc	2400
agacttaggc	acagcacgat	gcccaccacc	accagtgtgc	cgcacaaagg	cgtggcggtg	2460
gggtatgtgt	ctgaaaatga	gctcggggag	cgggcttgca	ccgctgacgc	atttgggaag	2520
cttaaggcag	cggcagaaga	agatgcaggc	agctgagttg	ttgtgttctg	ataagagtca	2580
gaggtaactc	ccgttgccgt	gctgttaacg	gtggagggca	gtgtagtctg	agcagtagtc	2640
gttgctgccg	cgcgcgccac	cgacataaat	agctgacaga	ctaacagact	gttcctttcc	2700
atgggtcttt	tctgcagtca	cccgggggat	ccttcgaacg	tagctctaga	ttgagtgcac	2760
gttactggcc	gaagccgctt	ggaataaggc	cggtgtgcgt	ttgtctatat	gttattttcc	2820
accatattgc	cgtctctttg	caatgtgagg	gcccggaaac	ctggccctgt	cttcttgacg	2880
agcattccta	ggggctcttt	ccctctcgcc	aaaggaatgc	aaggtctgtt	gaatgtogtg	2940
aaggaagcag	ttcctctgga	agcttcttga	agacaaacaa	cgtctgtagc	gaccttttgc	3000
aggcagcgga	acccccacc	tggcgacagg	tgcctctgcg	gccaaaagcc	acgtgtataa	3060
gatacacctg	caaaggcggc	acaaccccag	tgccacgttg	tgagttggat	agttgtggaa	3120
agagcctaat	ggctctcttc	aagcgtattc	aacaaggggc	tgaaggatgc	ccagaaggta	3180
ccccattgta	tgggatctga	tctggggcct	cggtgcacat	gctttacatg	tggttagtgc	3240
agggtaaaaa	aacgtctagg	ccccccgaac	cacggggacg	tggttttcc	ttgaaaaaca	3300
cgataataacc	atgggttaagt	gatattctact	agttgtgacc	ggcgccctagt	gttgacaatt	3360
aatcatcggc	atagtatatc	ggcatagtat	aatacgactc	actataggag	ggccaccatg	3420
tgcactacta	accttcttct	ctttcctaca	gctgagatca	ccggtaggag	ggccatcatg	3480

-14-

aaaaagcctg	aactcaccgc	gacgtctgtc	gcgaagtttc	tgatcgaaaa	gttcgacagc	3540
gtctccgacc	tgatgcagct	ctcggagggc	gaagaatctc	gtgctttcag	cttcgatgta	3600
ggagggcggtg	gatattgtcct	gcggttaa	agctgcgcgcg	atggtttcta	caaagatcgt	3660
tatgtttatc	ggcactttgc	atcggccgcg	ctcccgatc	cggagtgct	tgacattggg	3720
gaattcagcg	agagcctgac	ctattgcac	tcccgccgtg	cacagggtgt	cacgttgcaa	3780
gacctgcctg	aaaccgaact	gcccgcgtgt	ctgcaaccgc	tcgaggagct	catggatgcg	3840
atcgctgcgg	ccgatcttag	ccagacgagc	gggttcggcc	cattcggacc	gcaaggaa	3900
ggtcaataca	ctacatggcg	tgatttcata	tgcgcgattg	ctgatcccca	tgtgtatcac	3960
tggcaaactg	tgatggacga	caccgtcagt	gcgtccgtcg	cgcaggctct	cgatgagctg	4020
atgctttggg	ccgaggactg	ccccgaagtc	cggcacctcg	tgacgcggga	tttcggctcc	4080
aacaatgtcc	tgacggacaa	tggccgcata	acagcggcca	ttgactggag	cgaggcgatg	4140
ttcgggggatt	cccaatacga	ggtcgccaac	atcttcttct	ggaggccgtg	gttggcttgt	4200
atggagcagc	agacgcgcta	cttcgagcgg	aggcatccgg	agcttgacgg	atcgccgcgg	4260
ctccgggctg	atatgctccg	cattgggtctt	gaccaactct	atcagagctt	ggttgacggc	4320
aatttcgatg	atgcagcttg	ggcgccgggt	cgatgcgacg	caatcgtccg	atccggagcc	4380
gggactgtcg	ggcgctacaca	aatcgccgcg	agaagcgccg	cgtctgggac	cgatggctgt	4440
gtagaagtac	tcgccgatag	tggaaaccga	cgccccagca	ctcgtccgag	ggcaaaggaa	4500
tgagtcgaga	attcgctaga	gggcccattt	ctatagtgtc	acctaaatgc	tagagctcgc	4560
tgatcagcct	cgactgtgcc	ttctagtgtc	cagccatctg	ttgtttgccc	ctccccctg	4620
ccttccttga	ccctggaagg	tgccactccc	actgtccttt	cctaataaaa	tgaggaaatt	4680
gcacgcagct	gtctgagtag	gtgtcattct	attctggggg	gtgggggtggg	gcaggacagc	4740
aagggggagg	attgggaaga	caatagcagg	catgcgcagg	gcccatttgc	tcgagcggcc	4800
gcaataaaat	atctttatct	tcattacatc	tgtgtgttgg	ttttttgtgt	gaatcgtaac	4860
taacatacgc	tctccatcaa	aacaaaacga	aacaaaacaa	actagcaaaa	taggctgtcc	4920
ccagtgcagg	tgccaggtgcc	agaacatttc	tctatcgaag	gatctgcgat	cgctccgggtg	4980
cccgtcagtg	ggcagagcgc	acatcgccca	cagtcgccga	gaagttgggg	ggaggggtcg	5040
gcaattgaac	cggtgcctag	agaaggtggc	gcggggtaaa	ctgggaaagt	gatgtcgtgt	5100
actggctccg	cctttttccc	gaggggtggg	gagaaccgta	tataagtgca	gtagtcgccg	5160
tgaacgtttc	ttttcgcaac	gggtttgccc	ccagaacaca	gctgaagctt	cgaggggctc	5220
gcactctctc	ttcacgcgcg	cgccgcctta	cctgaggccg	ccatccacgc	cggttgagtc	5280
gcgttctgcc	gcctcccgcc	tgtggtgcct	cctgaactgc	gtccgcgcgtc	taggtaaagt	5340
taaagctcag	gtcgagaccg	ggcctttgtc	cggcgctccc	ttggagccta	cctagactca	5400
gccggctctc	cacgctttgc	ctgaccctgc	ttgctcaact	ctacgtcttt	gtttcgtttt	5460
ctgttctgcg	ccgttacaga	tccaagctgt	gaccggcgcc	tacgtaagtg	atatctacta	5520
gatttatcaa	aaagagtgtt	gacttgtgag	cgctcacaat	tgatacttag	attcatcgag	5580
agggacacgt	cgactactaa	ccttccttctc	tttcctacag	ctgagat		5627

<210> 28

<211> 553

<212> DNA

<213> Artificial Sequence

<220>

<223> pMG plasmid from InvivoGen: EMCV IRES sequence

<400> 28

aacgttactg	gccgaagccg	cttggaaataa	ggccgggtgtg	cgttttgtcta	tatgttatttt	60
tccaccatat	tgccgtcttt	tggcaatgtg	agggcccggga	aacctggccc	tgtcttcttg	120
acgagcattc	ctagggtctc	ttcccctctc	gccaaaaggaa	tgcaaggctc	gttgaatgtc	180
gtgaagggaag	cagttcctct	ggaagcttct	tgaagacaaa	caacgtctgt	agcgaccctt	240
tgcaggcagc	ggaacccccc	acctggcgac	aggtgcctct	gcggccaaaa	gccacgtgta	300
taagatacac	ctgcaaaggc	ggcacaaccc	cagtgcacacg	ttgtgagttg	gatagttgtg	360
gaaagagtca	aatggctctc	ctcaagcgta	ttcaacaagg	ggctgaagga	tgcccagaag	420
gtaccccatt	gtatgggata	tgatctgggg	cctcggtgca	catgctttac	gtgtgtttag	480
tcgagggttaa	aaaacgtcta	ggccccccga	accacgggga	cgtgggttttc	ctttgaaaaa	540
cacgatgata	ata					553

<210> 29

<211> 4692

<212> DNA

<213> Artificial Sequence

<220>

<223> pDSred1-N1 plasmid from Clontech

<400> 29

tagttatttaa	tagtaatcaa	ttacggggtc	attagttcat	agcccatata	tggagttccg	60
-------------	------------	------------	------------	------------	------------	----

-15-

cgttacataa	cttacggtaa	atggcccgc	tggetgacg	cccaacgacc	cccgccatt	120
gacgtcaata	atgacgtatg	ttcccatagt	aacgccaaata	gggactttcc	attgacgtca	180
atgggtggag	tatttacggt	aaactgccc	cttggcagta	catcaagtgt	atcatatgcc	240
aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgccagta	300
catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac	360
catggtgatg	cggtttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	420
atttccaagt	gtccaccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	480
ggactttcca	aaatgtcgta	acaactccgc	cccatgacg	caaattggcg	gtaggcgtgt	540
acggtgggag	gtctatataa	gcagagctgg	tttagtgaa	cgtcagatcc	gctagcgcta	600
ccggactcag	atctcgagct	caagcttcca	attctgcagt	cgacgggtacc	gcgggcccg	660
gatccaccgg	tcgccaccat	ggtgcgctcc	tccaagaacg	tcatacaagga	gttcattgcgc	720
ttcaaggtgc	gcattggagg	caccgtgaac	ggccacgagt	tcgagatcga	gggcgagggc	780
gagggccgccc	cctacgaggg	ccacaacacc	gtgaagctga	aggtgaccaa	gggcggcccc	840
ctgcccttcg	cctgggacat	cctgtccccc	cagttccagt	acggctccaa	ggtgtacgtg	900
aagcaccctcg	ccgacatccc	cgactacaag	aagctgtcct	ttcccagagg	cttcaagtgg	960
gagcgcgtga	tgaacttcga	ggacggcgcc	gtggtgacgg	tgaccacagga	ctcctccctg	1020
caggactcgt	gcttcattcta	caaggtgaag	ttcatcgagg	tgaacttccc	ctccgacggc	1080
cccgtaatgc	agaagaagac	catgggctgg	gaggcctcca	ccgagcgcct	gtacccccgc	1140
gacggcgtgc	tgaagggcga	gatccacaag	gccctgaagc	tgaaggacgg	cggccactac	1200
ctggtggagt	tcaagtccat	ctacatggcc	aagaagcccg	tcgagctgcc	cggctactac	1260
tacgtggagc	ccaagtggga	catcacctcc	cacaacgagg	actacaccat	cgtggagctag	1320
tacgagcgca	ccgagggccg	ccaccacctg	ttcctgtagc	ggccgcgact	ctagatcata	1380
atcagccata	ccacatttgt	agaggtttta	cttgctttta	aaaacctccc	acacctcccc	1440
ctgaacctga	aacataaaat	gaatgcaatt	gttgttgtta	acttgtttat	tgcagcttat	1500
aatggttaca	aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	tttttctactg	1560
cattctagtt	gtggtttgtc	caaaactcatc	aatgtatctt	aaggcgtaaa	ttgttaagctg	1620
taatattttg	ttaaaattcg	cgttaaabtt	ttgttaaatac	agctcatttt	ttaccaata	1680
ggccgaaatc	ggcaaaatcc	cttataaaatc	aaaagaatag	accgagatag	ggttgagtgt	1740
tgttccagtt	tggaaacaaga	gtccactatt	aaagaacgtg	gactccaacg	tcaaagggcg	1800
aaaaaccgtc	tatcaggggc	atggcccact	acgtgaacca	tcaccctaatt	caagtttttt	1860
ggggtcgagg	tgccgtaaaag	cactaaatcg	gaaccttaaa	gggagccccc	gatttagagc	1920
ttgacgggga	aagccggcga	acgtggcgag	aaaggaaggg	aagaaagcga	aaggagcggg	1980
cgctagggcg	ctggcaagtg	tagcggtcac	gctgcgcgta	accaccacac	ccgccgcgct	2040
taatgcgcgc	ctacagggcg	cgtcagggtg	cacttttctg	ggaaatgtgc	gcggaacccc	2100
tattgtttta	tttttctaaa	tacattcaaa	ttatgtatccg	ttcatgagac	aataaccctg	2160
ataaatgctt	caataatatt	gaaaaaggaa	gagtcctgag	gcggaagaaa	ccagctgtgg	2220
aatgtgtgtc	agttagggtg	tggaaagtcc	ccaggctccc	cagcaggcag	aagtatgcaa	2280
agcatgcata	tcaattagtc	agcaaccagg	tgtggaaagt	ccccaggctc	cccagcaggc	2340
agaagtatgc	aaagcatgca	tctcaattag	tcagcaacca	tagtcccgc	cctaactccg	2400
cccatcccg	ccctaactcc	gcccagttcc	gcccattctc	cgcccatgg	ctgactaatt	2460
ttttttattt	atgcagaggc	cgaggccgcc	tcggcctctg	agctattcca	gaagtagtga	2520
ggaggctttt	ttggaggcct	aggcttttgc	aaagatcgat	caagagacag	gatgaggatc	2580
gttttcgcatg	attgaacaag	atggattgca	cgcaggttct	ccggccgctt	gggtggagag	2640
gctattccgc	tatgactggg	cacaacagac	aatcggctgc	tctgatgccg	ccgtgttccg	2700
gctgtcagcg	caggggcgcc	cggttctttt	tgtaagagcc	gacctgtccg	gtgccctgaa	2760
tgaactgcaa	gacgaggcag	cgcggtatc	gtggctggcc	acgacggggc	ttccttgcgc	2820
agctgtgtct	gacgttgtca	ctgaagcggg	aagggaactg	ctgctattgg	gcgaagtgcc	2880
ggggcaggat	ctcctgtcat	ctcaccttgc	tcctgccgag	aaagtatcca	tcatggctga	2940
tgcaatgcgg	cggctgcata	cgcttgatcc	ggctacctgc	ccattcgacc	accaagcgaa	3000
acatcgcatc	gagcagcac	gtactcggat	ggaagccggg	cttgtcgatc	aggatgatct	3060
ggacgaagag	catcaggggc	tcgcgccagc	cgaactgttc	gccaggctca	aggcgagcat	3120
gcccgcaggg	gaggatctcg	tcgtgaccca	tggcgatgcc	tgcttgccga	atatcatggt	3180
ggaaaatggc	cgcttttctg	gattcatcga	ctgtggccgg	ctgggtgtgg	cggaccgcta	3240
tcaggacata	ccgttgatct	ccgtgatata	tgctgaagag	cttggcgggc	aatgggtcga	3300
ccgttctctc	gtgctttacg	gtatcgccgc	tcccgattcg	cagcgcatcg	ccttctatcg	3360
ccttcttgac	gagttcttct	gagcgggact	ctgggggttcg	aaatgaccga	ccaagcgacg	3420
cccaacctgc	catcacgaga	tttcgattcc	acgcgcgcct	tctatgaaag	gttgggcttc	3480
ggaatcgttt	tcggggacgc	cggtcggtg	atcctccagc	gcggggatct	catgctggag	3540
ttcttcggcc	gaggtctagg	gaggtctagg	gaaacacgga	aggagacaat	accggaagga	3600
acccgcgcta	tgacggcaat	aaaaagacag	aataaaacgc	acgggtgttg	gtcgtttgtt	3660
cataaacgcg	gggttcggct	ccagggtgg	cactctgtcg	ataccccacc	gagaccccat	3720
tggggccaat	acgcccgcgt	ttcttccctt	tccccacccc	accccccaag	ttcgggtgaa	3780
ggccacggcg	tcgcagccaa	cgtcggggcg	gcagccctcg	caatagcctc	aggttactca	3840
tatatacttt	agattgattt	aaaacttcat	tttttaattta	aaaggatcta	ggtgaagatc	3900
ctttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	3960
gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	tttttctgcg	cgtaatctgc	4020
tgcttgcaaa	caaaaaaacc	accgctacca	gcggtggttt	gtttgcgga	tcaagagcta	4080

-16-

ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	tactgtcctt	4140
ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcaccgcc	tacatacctc	4200
gctctgtctaa	tcctgttacc	agtggctgct	gccagtggcg	ataagtctgt	tcttaccggg	4260
ttggactcaa	gacgatagtt	accggataag	gcgcagcggt	cggtctgaac	gggggggttcg	4320
tgcacacagc	ccagcttgga	gcgaacgacc	tacaccgaac	tgagatacct	acagcgtgag	4380
ctatgagaaa	gcgccacgct	tcccgaaggg	agaaaggcgg	acaggtatcc	ggtaagcggc	4440
agggctggaa	caggagagcg	cacgaggag	cttccagggg	gaaacgcctg	gtatctttat	4500
agtcctgtcg	ggtttcgcc	cctctgactt	gagcgtcgat	ttttgtgatg	ctcgtcaggg	4560
ggcgagagcc	tatggaaaaa	cgccagcaac	gcggcctttt	tacggttcct	ggccttttgc	4620
tggccttttg	ctcacatggt	ctttcctgcg	ttatccctg	attctgtgga	taaccgtatt	4680
accgccatgc	at					4692

<210> 30

<211> 4257

<212> DNA

<213> Artificial Sequence

<220>

<223> pPur plasmid from Clontech

<400> 30

ctgtggaatg	tgtgtcagtt	aggggtgtgga	aagtccccag	gctccccagc	aggcagaagt	60
atgcaaagca	tgcattctcaa	ttagtcagca	accagggtgtg	gaaagtcccc	aggctcccca	120
gcaggcagaa	gtatgcaaag	catgcatctc	aattagtcag	caaccatagt	cccggccccta	180
actccgcccc	tcccggccct	aactccgccc	agttccgccc	attctccgcc	ccatggctga	240
ctaatttttt	ttattttatgc	agaggccgag	gcgcctcgg	cctctgagct	attccagaag	300
tagtgaggag	gctttttttg	aggcctaggc	ttttgcaaaa	agcttgcatg	cctgcaggtc	360
ggccgcacag	accggtgccg	ccaccatccc	ctgaccacag	cccctgacct	ctcacaagga	420
gacgaccttc	catgaccgag	tacaagccca	cggtgccgct	cgccaccgcc	gacgacgtcc	480
cccggggcgt	acgcaccctc	gccgcgcgct	tcgcccagta	ccccgccacg	cgccacaccg	540
tcgaccggga	ccgccacatc	gagcgggtca	ccgagctgca	agaactcttc	ctcacgcgcg	600
tcgggctcga	catcggaag	gtgtgggtcg	cggacgacgg	cgccgcgggtg	gcgggtctgga	660
ccacgcggga	gagcgtcgaa	gcgggggagg	tgttcgccga	gatcgggccc	cgcatggccg	720
agttgagcgg	ttcccggctg	gccgcgcagc	aacagatgga	aggcctcctg	gcgcccgcacc	780
ggcccaagga	gcccgcgtgg	ttcctggcca	ccgtcggcgt	ctcgcccagc	caccaggcca	840
agggctctgg	cagcgcgcgt	gtgctccccc	gagtggaggc	ggccgagcgc	gcccgggtgc	900
ccgccttcct	ggagacctcc	gcgccccgca	acctccccct	ctacgagcgg	ctcgggttca	960
ccgtcacccg	cgacgtcgag	gtgcccgaag	gaccgcgcac	ctgggtgcatg	acccgcaagc	1020
ccggtgcctg	acgcccggcc	cacgaccgcc	agcgcgccag	cgaaaggagc	gcacgacccc	1080
atggctccga	ccgaagccga	cccgggaggc	ccgcgcgacc	ccgcacccgc	ccccgaggcc	1140
caccgactct	agaggatcat	aatcagccat	accacatttg	tagaggtttt	acttgcttta	1200
aaaaacctcc	cacacctccc	cctgaacctg	aaacataaaa	tgaatgcaat	tgttgttgtt	1260
aacttggtta	ttgcagctta	taatggttac	aaataaagca	atagcatcac	aaatttcaca	1320
aataaagcat	tttttttcact	gcattctagt	tgtggtttgt	ccaaactcat	caatgtatct	1380
tatcatgtct	ggatccccag	gaagtccttc	tgtgtcctca	taaaccctaa	cctcctctac	1440
ttgagaggac	attccaatca	taggctgccc	atccaccctc	tgtgtcctcc	tggttaattag	1500
gtcacttaac	aaaaaggaaa	ttgggtaggg	gtttttcaca	gaccgctttc	taagggtaat	1560
tttaaaatat	ctgggaagtc	ccttccactg	ctgtgttcca	gaagtgttgg	taaacagccc	1620
acaaatgtca	acagcagaaa	catacaagct	gtcagctttg	cacaaggggc	caacaccctg	1680
ctcatcaaga	agcactgtgg	ttgctgtgtt	agtaatgtgc	aaaacaggag	gcacattttc	1740
cccacctgtg	taggtttccaa	aatatctagt	gtttttcattt	ttacttggat	caggaaccca	1800
gcactccact	ggataagcat	tatccttatc	caaaacagcc	ttgtggctcag	tggtcatctg	1860
ctgactgtca	actgtagcat	tttttggggg	tacagtttga	gcaggatatt	tggtcctgta	1920
gtttgctaac	acaccctgca	gctccaaagg	ttccccacca	acagcaaaaa	aatgaaaatt	1980
tgacccttga	atgggttttc	cagcaccatt	ttcatgagtt	ttttgtgtcc	ctgaatgcaa	2040
gtttaacata	gcagttaccc	caataacctc	agtttttaaca	gtaacagctt	cccacatcaa	2100
aatattttcca	caggttaagt	cctcatttaa	attaggcaaaa	ggaattcttg	aagacgaaag	2160
ggcctcgtga	tacgcctatt	tttataggtt	aatgtcatga	taataatggt	ttcttagacg	2220
tcaggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	tttgttttatt	tttctaaata	2280
cattcaataa	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	ataatattga	2340
aaaagggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	ttttgcccga	2400
ttttgccttc	ctgttttttc	tcaccagaaa	acgctgggtga	aagtaaaaga	tgctgaagat	2460
cagttggggc	cacgagtggt	ttacatcgaa	ctggactctca	acagcggtta	gatccttgag	2520
agttttcgcc	ccgaagaacg	ttttccaatg	atgagcactt	ttaaagttct	gctatgtggc	2580
gcggtattat	cccgtgttga	cgccgggcaa	gagcaactcg	gtcgccgcat	acactattct	2640
cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	tggcatgaca	2700
gtaagagaat	tatgcagtgc	tgccataacc	atgagtgata	acactgcggc	caacttactt	2760

-17-

ctgacaacga	tccggaggacc	gaaggagcta	accgcttttt	tgcacaacat	gggggatcat	2820
gtaactcgcc	ttgatcggtg	ggaaccggag	ctgaatgaag	ccataccaaa	cgacgagcgt	2880
gacaccacga	tgccctgcagc	aatggcaaca	acgttgcgca	aactattaac	tggcgaacta	2940
cttactctag	cttcccggca	acaattaata	gactggatgg	aggcggataa	agttgcagga	3000
ccacttctgc	gctcggccct	tccggctggc	tggtttattg	ctgataaaatc	tggagccggt	3060
gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggtaagcc	ctcccgtatc	3120
gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	acagatcgct	3180
gagataggtg	cctcactgat	taagcattgg	taactgtcag	accaagttta	ctcatatata	3240
cttttagattg	atttaaaact	tcatttttta	tttaaaagga	tctaggtgaa	gatccttttt	3300
gataatctca	tgacccaaaat	cccttaacgt	gagttttcgt	tccactgagc	gtcagacccc	3360
gtagaaaaga	tcaaaggatc	ttcttgagat	cctttttttc	tgcgcgtaat	ctgctgcttg	3420
caaacaaaaa	aaccaccgct	accagcgggtg	gtttgtttgc	cggatcaaga	gctaccaact	3480
ctttttccga	aggtaactgg	cttcagcaga	gcgcagatac	caaatactgt	ccttctagt	3540
tagccgtagt	taggocacca	cttcaagaac	tctgtagcac	cgccatacata	cctcgctctg	3600
ctaactcctgt	taccagtggc	tgctgccagt	ggcgataagt	cgtgtcttac	cgggttggac	3660
tcaagacgat	agttaccgga	taaggcgcag	cgttcgggct	gaacgggggg	ttcgtgcaca	3720
cagcccagct	tggagcgaac	gacctacacc	gaactgagat	acctacagcg	tgagctatga	3780
gaaagcgcca	cgcttcccga	agggagaaaag	gcggacaggt	atccggtaag	cggcagggtc	3840
ggaacaggag	agcgcacgag	ggagcttcca	gggggaaaacg	cctgggtatct	ttatagtcct	3900
gtcgggtttc	gccacctctg	acttgagcgt	cgatttttgt	gatgctcgtc	agggggggcg	3960
agcctatgga	aaaacgccag	caacgcggcc	tttttacgggt	tccctggcct	ttgctggcct	4020
tttgctcaca	tgttctttcc	tgcgttatcc	cctgattctg	tggataaccg	tattaccgcc	4080
tttgagttag	ctgataccgc	tgcgcgcagc	cgaacgaccg	agcgcagcga	gtcagtgagc	4140
gaggaagcgg	aagagcgcct	gatgcggtat	tttctcctta	cgcactctgtg	cggatatttca	4200
caccgcataat	ggtgcactct	cagtacaatc	tgctctgatg	ccgcatagtt	aagccag	4257

<210> 31

<211> 8136

<212> DNA

<213> Artificial Sequence

<220>

<223> pWE15 cosmid vector

<300>

<308> GenBank X65279

<309> 1995-04-14

<400> 31

ctatagtgag	tcgattattg	cggccgcgaa	ttcttgaaga	cgaaagggcc	tcgatgatac	60
cctatttttta	taggttaatg	tcattgataat	aatgggtttct	tagacgtcag	gtggcacttt	120
tccggggaaat	gtgcgcggaa	cccctatttg	tttatttttc	ttaaatacatt	caaatatgta	180
tccgctcatg	agacaataac	cctgataaat	gcttcaataa	tattgaaaaa	ggaagagtat	240
gagtattcaa	catttccgtg	tgcgccctat	tccctttttt	gcggcatttt	gcttccctgtt	300
tttgctcacc	cagaaacgct	ggtgaaagta	aaagatgctg	aagatcagtt	gggtgcacga	360
gtgggttaca	tgaactgga	tctcaacagc	ggtaagatcc	ttgagagttt	tgcgcccgaa	420
gaacgttttc	caatgatgag	cactttttaa	gttctgctat	gtggcgcggt	attatcccgt	480
gttgacgcgc	ggcaagagca	actcggctgc	cgcatacact	attctcagaa	tgacttgggt	540
gagtactcac	cagtcacaga	aaagcatctt	acggatggca	tgacagtaag	agaattatgc	600
agtgtgcca	taaccatgag	tgataaacact	gcggccaact	tacttctgac	aacgatcgga	660
ggaccgaagg	agctaaccgc	ttttttgcac	aacatggggg	atcatgtaac	tgccttgat	720
cgttgggaac	cggagctgaa	tgaagccata	ccaaacgacg	agcgtgacac	cacgatgcct	780
gcagcaatgg	caacaacggt	gcgcaaacta	ttactggcgc	aactacttac	tctagcttcc	840
cggcaacaat	taatagactg	gatggaggcg	gataaagttg	caggaccact	tctgcgctcg	900
gcccttccgg	ctggctgggt	tattgctgat	aaatctggag	cgggtgagcg	tgggtctcgc	960
ggtatcattg	cagcactggg	gccagatggg	aagcctcccc	gtatcgtagt	tatctacacg	1020
acggggagtc	aggcaactat	ggatgaacga	aatagacaga	tcgctgagat	aggtgcctca	1080
ctgattaaac	attggtaac	gtcagaccga	gttactcat	atatacttta	gattgattta	1140
aaacttcatt	tttaatttaa	aaggatctag	gtgaagatcc	tttttgataa	tctcatgacc	1200
aaaatccctt	aacgtgagtt	ttcgttccac	tgagcgtcag	accccgtaga	aaagatcaaa	1260
ggatcttctt	gagatccctt	ttttctgcgc	gtaatctgct	gcttgcaaac	aaaaaaacca	1320
ccgctaccag	cgggtggttg	tttgccggat	caagagctac	caactctttt	tccgaaggta	1380
actggcttca	gcagagcgca	gataccaaat	actgtctctc	tagtgtagcc	gtagttaggc	1440
caccacttca	agaactctgt	agcaccgcct	acatacctcg	ctctgctaata	cctgttacca	1500
gtggctgctg	ccagtggcga	taagtcgtgt	cttaccgggt	tggactcaag	acgatagtta	1560
cgggataaag	cgcagcggtc	gggctgaacg	gggggttcgt	gcacacagcc	cagcttggag	1620
cgaacgacct	acaccgaact	gagataccta	cagcgtgagc	tatgagaaag	cgccacgctt	1680

-18-

cgaagggag	aaagggcgac	aggtatccgg	taagcggcag	ggtcggaaca	ggagagcgca	1740
cgagggagct	tccaggggga	aacgcctggg	atctttatag	tctgtcggg	gtttcgccac	1800
ctctgacttg	agcgtcgatt	tttgtgatgc	tcgtcagggg	ggcggagcct	atggaaaaac	1860
gccagcaacg	cggccttttt	acgggttcctg	gccttttgct	ggccttttgc	tcacatgttc	1920
tttctgctg	tatcccttga	ttctgtggat	aaccgtatta	ccgccttttga	gtgagctgat	1980
accgctcgcc	gcagccgaac	gaccgagcgc	agcgagtcag	tgagcgagga	agcggaagag	2040
cgctgacttc	cgcgtttcca	gactttacga	aacacggaaa	ccgaagacca	ttcatgttgt	2100
tgctcaggtc	gcagacgttt	tgcagcagca	gtcgcttcac	gttcgctcgc	gtatcgggtga	2160
ttcattctgc	taaccagtaa	ggcaaccccc	ccagcctagc	cggtgcctca	acgacaggag	2220
cacgatcatg	cgcacccgtc	agatccagac	atgataagat	acattgatga	gtttggacaa	2280
accacaacta	gaatgcagtg	aaaaaaatgc	tttatttctg	aaatttgtga	tgctattgct	2340
ttatttgtaa	ccattataag	ctgcaataaaa	caagttaaca	acaacaattg	cattcattttt	2400
atgtttcagg	ttcaggggga	ggtgtgggag	gttttttaaa	gcaagtaaaa	cctctacaaa	2460
tgtggtatgg	ctgattatga	tctctagtca	aggcactata	catcaaatat	tccttattaa	2520
cccttttaca	aattaaaaag	ctaaaaggtag	acaatttttg	agcatagtta	ttaatagcag	2580
acactctatg	cctgtgtgga	gtaagaaaaa	acagtatgtt	atgattataa	ctgttatgoc	2640
tacttataaa	ggttacagaa	tattttttcca	taattttctt	gtatagcagt	gcagcttttt	2700
cctttgtggt	gtaaatagca	aagcaagcaa	gagttctatt	actaaacaca	gcatgactca	2760
aaaaacttag	caattctgaa	ggaaagtcct	tggggctctc	tacctttctc	ttcttttttg	2820
gaggagtaga	atgttgagag	tcagcagtag	cctcatcctc	actagatggc	attcttctctg	2880
agcaaaacag	gttttctcca	ttaaaggcat	tccaccactg	ctccattcca	tcagttccat	2940
aggttggaat	ctaaaatata	caaacaatta	gaatcagtag	tttaacacat	tatacactta	3000
aaaattttat	atttacctta	gagcttttaa	tctctgtagg	tagtttgtcc	aattatgtca	3060
caccacagaa	gtaaggttcc	ttcacaagaa	tccggaccac	agcggccatc	gtgcctcccc	3120
actcctgcag	ttcgggggca	tggatgcgcg	gatagccgct	gctgggttcc	tggatgcgca	3180
cggattttga	ctgcgcgtag	aactcgcgag	gtcgctccagc	ctcaggcagc	agctgaacca	3240
actcgcgagg	ggatcgagcc	cgggggtgggc	gaagaactcc	agcatgagat	ccccgcgctg	3300
gaggatcatc	cagccggcgt	cccggaaaaac	gattccgaag	cccaaccttt	catagaaggc	3360
ggcgggtgaa	tcgaaatctc	gtgatggcag	gttggggctc	gcttgggtcg	tcatttcgaa	3420
ccccagagtc	ccgctcagaa	gaactcgtca	agaaggcgat	agaaggcgat	gcgctgcgaa	3480
tcggggagcgg	cgataccgta	aagcacgagg	aagcggtcag	cccattcgcc	gccaagctct	3540
tcagcaatat	cacgggtagc	caacgcctatg	tcctgatagc	ggtcgcgccac	accagcccg	3600
ccacagtcga	tgaatccaga	aaagcggcca	ttttccacca	tgatattcgg	caagcaggca	3660
tcgccatggg	tcacgacgag	atcctcgccc	tcgggatgcg	cgcttgagc	ctggcgaaac	3720
gttcggctgg	cgcgagcccc	ttgatgctctt	cgctcagatc	atcctgatcg	acaagaccgg	3780
cttccatccg	agtacgtgct	cgctcgatgc	gatgtttcgc	ttgggtggctg	aatgggcagg	3840
tagccggatc	aagcgtatgc	agccgcgcga	ttgcatcagc	catgatggat	actttctcgg	3900
caggagcaag	gtgagatgac	aggagatcct	gccccggcac	ttcgcccaat	agcagccagt	3960
cccttcccg	ccctcgagca	acgtcgagca	cagctcgcca	aggaacgccc	gtcgtgcgca	4020
gccacgatag	ccgcgctgcc	tcgtcctgca	gttcattcag	ggcaccggac	aggctggctc	4080
tgacaaaaag	aaccggggcg	ccctgcgctg	acagccggaa	cacggcggca	tcagagcagc	4140
cgattgtctg	ttgtgcccag	tcatagccga	atagcctctc	cacccaagcg	gcccggagaac	4200
ctgcgtgcaa	tccatcttgt	tcaatcatgc	gaaacgatcc	tcattcctgtc	tcttgatcag	4260
atcttgatcc	cctgcgccat	cagatccttg	cgggcaagaa	agccatccag	ttacttttgc	4320
agggcttccc	aaccttacca	gagggcgccc	cagctggcaa	ttccgggttcg	cttgctgtcc	4380
ataaaaccgc	ccagtctagc	tatcgccatg	taagccact	gcaagctacc	tgctttctct	4440
ttgcgcttgc	gttttccctt	gtccagatag	cccagtagct	gacattcatc	cggggctcagc	4500
accgtttctg	cggactggct	ttctacgtgt	tcgccttctt	ttagcagccc	ttgcgccttg	4560
agtgccttgc	gcagcgtgaa	agcttttttg	aaaagcctag	gcctccaaaa	aagcctcttc	4620
actacttctg	gaatagctca	gagggcgagg	cggcctaaat	aaaaaaaatt	agtcagccat	4680
ggggcggaga	atggggcgga	ctggggcgag	ttagggcgcg	gatgggcgga	gttaggggag	4740
ggactatggt	tgctgactaa	ttgagatgca	tgctttgcat	acttctgcct	gctggggagc	4800
ctggggactt	tccacaccctg	gttgctgact	aattgagatg	catgctttgc	atacttctgc	4860
ctgctgggga	gcctggggac	tttccacacc	ctaactgaca	cacattccac	agccggatct	4920
gcaggaccca	acgctgcccc	agatgcgcgc	cgtgcggctg	ctggagatgg	cggacgcgat	4980
ggatatgttc	tgccaagggg	tggttttgcc	attcacagtt	ctccgcaaga	attgattggc	5040
tccaattctt	ggagtgggtga	atccgttagc	gaggtgccc	cggcttccat	tcaggtcgag	5100
gtggcccgcc	tccatgcacc	gcgacgcaac	gcggggagcg	agacaaagga	tagggcgccg	5160
cctacaatcc	atgccaaacc	gttccatgtg	ctcgccgagg	cgcataaaatc	gccgtgacga	5220
tcagcggctc	aatgatcgaa	gttaggctgg	taagagccgc	gagcgatcct	tgaagctgtc	5280
cctgatggtc	gtcatctacc	tgccctggaca	gcattggctg	caacgcggca	tcccgatgcc	5340
gccggaagcg	agaagaatca	taatggggaa	ggccattccag	cctcgcgctg	cgaacgccag	5400
caagacgtag	ccagcgcgt	cgggccgcca	tgccggcgat	aatggcctgc	ttctcgccga	5460
aacgtttggt	ggcgggacca	gtgacgaagg	cttgagcgag	ggcgtgcaag	attccgaata	5520
ccgcaagcga	caggccgatc	atcgctcgcc	tccagcgaaa	gcggctcctg	ccgaaaatga	5580
cccagagcgc	tgccggcacc	tgtcctacga	gttgcatgat	aaagaagaca	gtcataagtg	5640
cggcgacgat	agtcatgccc	cgcgcccacc	ggaaggagct	gactgggttg	aaggctctca	5700

-19-

agggcatcg	tcgacgctct	cccttatg	actcctgcat	taggaagcag	cccagtagta	5760
ggttgaggg	gttgagcacc	gccgcccga	ggaatggtgc	atgcaaggag	atggcgccca	5820
acagtcccc	ggccacgggc	ctgccacat	acccacgccc	aaacaagcgc	tcatgagccc	5880
gaagtggcg	gcccgatctt	ccccatcgg	gatgtcggcg	atataggcgc	cagcaaccgc	5940
acctgtggcg	ccggtgatgc	cggccacgat	gcgtccggcg	tagaggatct	tggcagtcac	6000
agcatgcgca	tatccatgct	tcgaccatgc	gctcacaag	taggtgaatg	cgcaatgtag	6060
taccacacac	gtcatcgctt	tccactgctc	tcgcaataa	agatggaaaa	tcaatctcat	6120
ggtaatagtc	catgaaaatc	cttgtattca	taaatcctcc	aggtagctat	atgcaaattg	6180
aaacaaaaga	gatggtgatc	tttctaagag	atgatggaat	ctcccttcag	tatcccgatg	6240
gtcaatgcgc	tggatatggg	atagatggga	atatgctgat	ttttatggga	cagagttgcg	6300
aactgttccc	aactaaaatc	atcttgccac	atcagcgcac	tacgaacttt	accacaaaat	6360
agtcaggtaa	tgaatcctga	tataaagaca	ggttgataaa	tcagtcttct	acgcgcatcg	6420
cacgcgcaca	ccgtagaaag	tctttcagtt	gtgagcctgg	gcaaaccggt	aactttcggc	6480
ggctttgctg	tcgcacaggc	tcacgtctaa	aaggaaataa	atcatgggtc	ataaaattat	6540
cacgttgtcc	ggcgcgccga	cggatgttct	gtatgcgctg	tttttcogtg	gcgcgttgct	6600
gtctggtgat	ctgccttcta	aatctggcac	agccgaattg	cgcgagcttg	gttttgctga	6660
aaccagacac	acagcaactg	aataccagaa	agaaaatcac	tttacctttc	tgacatcaga	6720
agggcagaaa	tttgccgttg	aacacctggt	caatacgcgt	tttggtgagc	agcaatattg	6780
cgcttcgatg	acgcttgggc	ttgagattga	tacctctgct	gcacaaaagg	caatcgacga	6840
gctggaccag	cgcattcgtg	acaccgtctc	cttcgaactt	attcgcaatg	gagtgctcat	6900
catcaaggag	gcgcgtatcg	caaatgggtc	tatccacgca	gcggcaatcg	aaacacctca	6960
gccggtgacc	aatatctaca	acatcagcct	tggtatccag	cgtgatgagc	cagcgcagaa	7020
caaggtaacc	gtcagtgccg	ataagttcaa	agttaaacct	ggtggtgata	ccaacattga	7080
aacgttgatc	gaaaacgcgc	tgaaaaacgc	tgctgaatgt	gcggcgctgg	atgtcacaaa	7140
gcaaatggga	gcagacaaga	aagcgaatga	tgaactggct	tcctatgtcc	gcacggccat	7200
catgatggaa	tgtttccccc	gtgggtgttat	ctggcagcag	tgccgtcgat	agtatgcaat	7260
tgataattat	tatcattttg	gggtcctttc	cggcgatccg	ccttgttacg	gggcggcgac	7320
ctcgcggtt	ttcgctattt	atgaaaattt	tccggtttaa	ggcggtttcc	ttcttcttcg	7380
tcataactta	atgtttttat	ttaaaatacc	ctctgaaaag	aaaggaaaac	acaggtgctg	7440
aaagcgaact	ttttggcctc	tgtcgtttcc	tttctctggt	tttgtccgtg	gaatgaacaa	7500
tggaagtcaa	caaaaagcag	ctggctgaca	ttttcgggtg	gagtatccgt	accattcaga	7560
actggcagga	acaggggaatg	cccgttctgc	gaggcggtgg	caagggtaat	gaggtgcttt	7620
atgactctgc	cgccgtcata	aaatggtag	ccgaaaggga	tgctgaaatt	gagaacgaaa	7680
agctgcgcgc	ggaggttgaa	gaactgcggc	aggccagcga	ggcagatcca	caggacgggt	7740
gtggtcgcca	gtcgatagtg	gtcgatagtg	gtcccaagta	gcgaagcgag	caggactggg	7800
cggcgccaaa	gcggctcgac	agtgtctccg	gaacgggtgc	gcatagaaat	tgcatcaacg	7860
catatagcgc	tagcagcacg	ccatagtgac	tgccgatgct	gtcgggaatg	acgatatccc	7920
gcaagaggcc	cggcagtagc	ggcataacca	agcctatgcc	tacagcatcc	agggtgacgg	7980
tgccaggagt	gcagatgagc	cgattgttag	atttcataca	cgggtgcctga	ctgcgttagc	8040
aatttaactg	tgataaacta	cgcattaaag	gcttatcgat	gataagcggg	caaacatgag	8100
aattcgcggc	cgcaattaac	cctcactaaa	ggatcc			8136

<210> 32

<211> 2713

<212> DNA

<213> Artificial Sequence

<220>

<223> pNEB193 plasmid

<400> 32

tcgcgcgttt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacgggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccc	tcaggggcgc	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accatatg	gtgtgaaata	cgcacagat	gcgttaaggag	aaaataccgc	atcaggcgcc	240
attcgccatt	caggctgcgc	aactgttggg	aagggcgatc	ggtgcggggc	tcttcgctat	300
tacgccagct	ggcgaaaggg	ggatgtgctg	caaggcgatt	aagttgggtg	acgccagggt	360
tttcccagtc	acgacgttgt	aaaacgacgc	ccagtgaaat	cgagctcggt	acccgggggc	420
gcgcgggac	cttaattaa	tctagagtcg	actgtttaaa	cctgcaggca	tgcaagcttg	480
gcgtaatcat	ggatcatagc	gtttcctgtg	tgaaattgtt	atccgctcac	aattccacac	540
aacatacgag	ccggaagcat	aaagtgtaaa	gcctgggggtg	cctaattgag	gagctaactc	600
acattaattg	cgttgcgctc	actgcccgcg	ttccagtcgg	gaaacctgtc	gtgccagctg	660
cattcaatgaa	tcggccaacg	cgcggggaga	ggcggtttgc	gtattggggc	ctcttcgctc	720
tcctcgctca	ctgactcgct	gcgctcggtc	gttcgggtgc	ggcgagcggg	atcagctcac	780
tcaaaggcgg	taatacgggt	atccacagaa	tcaggggata	acgcaggaaa	gaacatgtga	840
gcaaaaggcc	agcaaaaggc	caggaaccgt	aaaaaggccg	cgttgctggc	gtttttccat	900
aggctccgcc	cccctgacga	gcatacaaaa	aatcgacgct	caagtacagag	gtggcgaaac	960

-20-

ccgacaggac	tataaagata	ccaggcggtt	ccccctggaa	gctccctcgt	gcgctctcct	1020
gttccgaccc	tgccgcttac	cggatacctg	tccgcctttc	tcccttcggg	aagcgtggcg	1080
ctttctcata	gctcacgctg	taggtatctc	agttcggtgt	aggtcgttcg	ctccaagctg	1140
ggctgtgtgc	acgaaccccc	cgttcagccc	gaccgctgcg	ccttatccgg	taactatcgt	1200
cttgagtcca	acccggtaag	acacgactta	tcgccactgg	cagcagccac	tggtaacagg	1260
attagcagag	cgaggatatg	aggcgggtgt	acagagttct	tgaagtgggtg	gcctaactac	1320
ggctacacta	gaaggacagt	atttgggtatc	tcgctctctgc	tgaagccagt	taccttcgga	1380
aaaagagttg	gtagctcttg	atccggcaaa	caaaccaccg	ctggtagcgg	tggttttttt	1440
gtttgcaagc	agcagattac	gcgcagaaaa	aaaggatctc	aagaagatcc	tttgatcttt	1500
tctacggggt	ctgacgctca	gtggaacgaa	aactcacgtt	aagggatattt	ggtcatgaga	1560
ttatcaaaaa	ggatcttcac	ctagatcctt	ttaaattaaa	aatgaagttt	taaatcaatc	1620
taaagtatat	atgagtaaac	ttggtctgac	agttaccaat	gcttaatcag	tgaggcacct	1680
atctcagcga	tctgtctatt	tcgttcaccc	atagttgcct	gactccccgt	cgtgtagata	1740
actacgatac	gggagggctt	accatctggc	cccagtgctg	caatgatacc	gcgagaccca	1800
cgctcaccgg	ctccagattt	atcagcaata	aaccagccag	ccggaagggc	cgagcgcaga	1860
agtggtcctg	caactttatc	cgctccatc	cagtctatta	attggtgccc	ggaagctaga	1920
gtaagtagtt	cgccagttaa	tagtttgctc	aacgttggtg	ccattgctac	aggcatcgtg	1980
gtgtcacgct	cgtcgtttgg	tatggcttca	ttcagctccg	gttccaacg	atcaaggcga	2040
gttacatgat	cccccatggt	gtgcaaaaa	gcggttagct	ccttcgggtc	tccgatcggt	2100
gtcagaagta	agttggccgc	agtgttatca	ctcatgggta	tggcagcact	gcataattct	2160
cttactgtca	tgccatccgt	aagatgcttt	tctgtgactg	gtgagtactc	aaccaagtca	2220
ttctgagaat	agtgtatgcg	gcgaccgagt	tgctcttgcc	cggcgtcaat	acgggataat	2280
accgcgcac	atagcagaac	tttaaaagtg	ctcatcattg	gaaaacgttc	ttcggggcga	2340
aaactctcaa	ggatcttacc	gctggtgaga	tccagttcga	tgtaacccac	tcgtgcaccc	2400
aactgatctt	cagcatcttt	tactttcacc	agcgtttctg	ggtgagcaaa	aacaggaagg	2460
caaaatgccg	caaaaaaggg	aataagggcg	acacggaaat	gttgaatact	catactcttc	2520
ctttttcaat	attattgaag	catttatcag	ggttattgtc	tcatgagcgg	atacatattt	2580
gaatgtattt	agaaaaataa	acaaataggg	gttccgcgca	catttccccg	aaaagtgccg	2640
cctgacgtct	aagaaaccat	tattatcatg	acattaacct	ataaaaaatag	gcgtatcacg	2700
aggccctttc	gtc					2713

<210> 33

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> attP

<400> 33

cagctttttt atactaagtt g

21

<210> 34

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> attB

<400> 34

ctgctttttt atactaactt g

21

<210> 35

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> attL

<400> 35

ctgctttttt atactaagtt g

21

<210> 36

<211> 21

<212> DNA

-21-

<213> Artificial Sequence

<220>

<223> attR

<400> 36

cagctttttt atactaactt g

21

<210> 37

<211> 1071

<212> DNA

<213> Artificial Sequence

<220>

<223> Integrase E174R

<221> CDS

<222> (1)...(1071)

<223> Nucleotide sequence encoding Integrase E147R

<400> 37

atg gga aga agg cga agt cat gag cgc cgg gat tta ccc cct aac ctt	48
Met Gly Arg Arg Arg Ser His Glu Arg Arg Asp Leu Pro Pro Asn Leu	
1 5 10 15	
tat ata aga aac aat gga tat tac tgc tac agg gac cca agg acg ggt	96
Tyr Ile Arg Asn Asn Gly Tyr Tyr Cys Tyr Arg Asp Pro Arg Thr Gly	
20 25 30	
aaa gag ttt gga tta ggc aga gac agg cga atc gca atc act gaa gct	144
Lys Glu Phe Gly Leu Gly Arg Asp Arg Arg Ile Ala Ile Thr Glu Ala	
35 40 45	
ata cag gcc aac att gag tta ttt tca gga cac aaa cac aag cct ctg	192
Ile Gln Ala Asn Ile Glu Leu Phe Ser Gly His Lys His Lys Pro Leu	
50 55 60	
aca gcg aga atc aac agt gat aat tcc gtt acg tta cat tca tgg ctt	240
Thr Ala Arg Ile Asn Ser Asp Asn Ser Val Thr Leu His Ser Trp Leu	
65 70 75 80	
gat cgc tac gaa aaa atc ctg gcc agc aga gga atc aag cag aag aca	288
Asp Arg Tyr Glu Lys Ile Leu Ala Ser Arg Gly Ile Lys Gln Lys Thr	
85 90 95	
ctc ata aat tac atg agc aaa att aaa gca ata agg agg ggt ctg cct	336
Leu Ile Asn Tyr Met Ser Lys Ile Lys Ala Ile Arg Arg Gly Leu Pro	
100 105 110	
gat gct cca ctt gaa gac atc acc aca aaa gaa att gcg gca atg ctc	384
Asp Ala Pro Leu Glu Asp Ile Thr Thr Lys Glu Ile Ala Ala Met Leu	
115 120 125	
aat gga tac ata gac gag ggc aag gcg gcg tca gcc aag tta atc aga	432
Asn Gly Tyr Ile Asp Glu Gly Lys Ala Ala Ser Ala Lys Leu Ile Arg	
130 135 140	
tca aca ctg agc gat gca ttc cga gag gca ata gct gaa ggc cat ata	480
Ser Thr Leu Ser Asp Ala Phe Arg Glu Ala Ile Ala Glu Gly His Ile	
145 150 155 160	
aca aca aac cat gtc gct gcc act cgc gca gca aaa tct aga gta agg	528
Thr Thr Asn His Val Ala Ala Thr Arg Ala Ala Lys Ser Arg Val Arg	
165 170 175	
aga tca aga ctt acg gct gac gaa tac ctg aaa att tat caa gca gca	576
Arg Ser Arg Leu Thr Ala Asp Glu Tyr Leu Lys Ile Tyr Gln Ala Ala	

-22-

180	185	190	
gaa tca tca cca tgt tgg ctc aga ctt gca atg gaa ctg gct gtt gtt Glu Ser Ser Pro Cys Trp Leu Arg Leu Ala Met Glu Leu Ala Val Val 195 200 205			624
acc ggg caa cga gtt ggt gat tta tgc gaa atg aag tgg tct gat atc Thr Gly Gln Arg Val Gly Asp Leu Cys Glu Met Lys Trp Ser Asp Ile 210 215 220			672
gta gat gga tat ctt tat gtc gag caa agc aaa aca ggc gta aaa att Val Asp Gly Tyr Leu Tyr Val Glu Gln Ser Lys Thr Gly Val Lys Ile 225 230 235			720
gcc atc cca aca gca ttg cat att gat gct ctc gga ata tca atg aag Ala Ile Pro Thr Ala Leu His Ile Asp Ala Leu Gly Ile Ser Met Lys 245 250 255			768
gaa aca ctt gat aaa tgc aaa gag att ctt ggc gga gaa acc ata att Glu Thr Leu Asp Lys Cys Lys Glu Ile Leu Gly Gly Glu Thr Ile Ile 260 265 270			816
gca tct act cgt cgc gaa ccg ctt tca tcc ggc aca gta tca agg tat Ala Ser Thr Arg Arg Glu Pro Leu Ser Ser Gly Thr Val Ser Arg Tyr 275 280 285			864
ttt atg cgc gca cga aaa gca tca ggt ctt tcc ttc gaa ggg gat ccg Phe Met Arg Ala Arg Lys Ala Ser Gly Leu Ser Phe Glu Gly Asp Pro 290 295 300			912
cct acc ttt cac gag ttg cgc agt ttg tct gca aga ctc tat gag aag Pro Thr Phe His Glu Leu Arg Ser Leu Ser Ala Arg Leu Tyr Glu Lys 305 310 315			960
cag ata agc gat aag ttt gct caa cat ctt ctc ggg cat aag tcg gac Gln Ile Ser Asp Lys Phe Ala Gln His Leu Leu Gly His Lys Ser Asp 325 330 335			1008
acc atg gca tca cag tat cgt gat gac aga ggc agg gag tgg gac aaa Thr Met Ala Ser Gln Tyr Arg Asp Asp Arg Gly Arg Glu Trp Asp Lys 340 345 350			1056
att gaa atc aaa taa Ile Glu Ile Lys *			1071
355			
<210> 38			
<211> 356			
<212> PRT			
<213> Artificial Sequence			
<220>			
<223> Integrase E147R			
<400> 38			
Met Gly Arg Arg Arg Ser His Glu Arg Arg Asp Leu Pro Pro Asn Leu 1 5 10 15			
Tyr Ile Arg Asn Asn Gly Tyr Tyr Cys Tyr Arg Asp Pro Arg Thr Gly 20 25 30			
Lys Glu Phe Gly Leu Gly Arg Asp Arg Arg Ile Ala Ile Thr Glu Ala 35 40 45			
Ile Gln Ala Asn Ile Glu Leu Phe Ser Gly His Lys His Lys Pro Leu 50 55 60			
Thr Ala Arg Ile Asn Ser Asp Asn Ser Val Thr Leu His Ser Trp Leu 65 70 75 80			
Asp Arg Tyr Glu Lys Ile Leu Ala Ser Arg Gly Ile Lys Gln Lys Thr			

-23-

[illegible]

```
<210> 39
<211> 876
<212> DNA
<213> Discosoma species
```

```
<220>
<221> CDS
<222> (45)...(737)
<223> Nucleotide sequence encoding red fluorescent
      protein (FP593)
```

<300>
<308> GenBank AF272711
<309> 2000-09-26

```
<400> 39
agtttcagcc agtgacaggg tgagctgcc a ggtattctaa caag atg agt tgt tcc 56
                               Met Ser Cys Ser
                               1
```

aag aat gtg atc aag gag ttc atg agg ttc aag gtt cgt atg gaa gga 104
Lys Asn Val Ile Lys Glu Phe Met Arg Phe Lys Val Arg Met Glu Gly
5 10 15 20

acg gtc aat ggg cac gag ttt gaa ata aaa ggc gaa ggt gaa ggg agg 152
Thr Val Asn Gly His Glu Phe Glu Ile Lys Gly Glu Gly Glu Gly Arg
25 30 35

cct tac gaa ggt cac tgt tcc gta aag ctt atg gta acc aag ggt gga 200
Pro Tyr Glu Gly His Cys Ser Val Lys Leu Met Val Thr Lys Gly Gly

-24-

40										45										50										
cct	ttg	cca	ttt	gct	ttt	gat	att	ttg	tca	cca	caa	ttt	cag	tat	gga	248														
Pro	Leu	Pro	Phe	Ala	Phe	Asp	Ile	Leu	Ser	Pro	Gln	Phe	Gln	Tyr	Gly															
		55					60					65																		
agc	aag	gta	tat	gtc	aaa	cac	cct	gcc	gac	ata	cca	gac	tat	aaa	aag	296														
Ser	Lys	Val	Tyr	Val	Lys	His	Pro	Ala	Asp	Ile	Pro	Asp	Tyr	Lys	Lys															
	70					75					80																			
ctg	tca	ttt	cct	gag	gga	ttt	aaa	tgg	gaa	agg	gtc	atg	aac	ttt	gaa	344														
Leu	Ser	Phe	Pro	Glu	Gly	Phe	Lys	Trp	Glu	Arg	Val	Met	Asn	Phe	Glu															
	85				90					95					100															
gac	ggt	ggc	gtg	gtt	act	gta	tcc	caa	gat	tcc	agt	ttg	aaa	gac	ggc	392														
Asp	Gly	Gly	Val	Val	Thr	Val	Ser	Gln	Asp	Ser	Ser	Leu	Lys	Asp	Gly															
				105					110					115																
tgt	ttc	atc	tac	gag	gtc	aag	ttc	att	ggg	gtg	aac	ttt	cct	tct	gat	440														
Cys	Phe	Ile	Tyr	Glu	Val	Lys	Phe	Ile	Gly	Val	Asn	Phe	Pro	Ser	Asp															
			120					125					130																	
gga	cct	gtt	atg	cag	agg	agg	aca	cgg	ggc	tgg	gaa	gcc	agc	tct	gag	488														
Gly	Pro	Val	Met	Gln	Arg	Arg	Thr	Arg	Gly	Trp	Glu	Ala	Ser	Ser	Glu															
		135					140					145																		
cgt	ttg	tat	cct	cgt	gat	ggg	gtg	ctg	aaa	gga	gac	atc	cat	atg	gct	536														
Arg	Leu	Tyr	Pro	Arg	Asp	Gly	Val	Leu	Lys	Gly	Asp	Ile	His	Met	Ala															
	150					155					160																			
ctg	agg	ctg	gaa	gga	ggc	ggc	cat	tac	ctc	gtt	gaa	ttc	aaa	agt	att	584														
Leu	Arg	Leu	Glu	Gly	Gly	Gly	His	Tyr	Leu	Val	Glu	Phe	Lys	Ser	Ile															
	165					170				175					180															
tac	atg	gta	aag	aag	cct	tca	gtg	cag	ttg	cca	ggc	tac	tat	tat	gtt	632														
Tyr	Met	Val	Lys	Lys	Pro	Ser	Val	Gln	Leu	Pro	Gly	Tyr	Tyr	Tyr	Val															
				185					190					195																
gac	tcc	aaa	ctg	gat	atg	acg	agc	cac	aac	gaa	gat	tac	aca	gtc	gtt	680														
Asp	Ser	Lys	Leu	Asp	Met	Thr	Ser	His	Asn	Glu	Asp	Tyr	Thr	Val	Val															
			200					205					210																	
gag	cag	tat	gaa	aaa	acc	cag	gga	cgc	cac	cat	ccg	ttc	att	aag	cct	728														
Glu	Gln	Tyr	Glu	Lys	Thr	Gln	Gly	Arg	His	His	Pro	Phe	Ile	Lys	Pro															
		215					220					225																		
ctg	cag	tga	actcggctca	gtcatggatt	agcggtaatg	gccacaaaag										777														
Leu	Gln	*																												
	230																													
gcacgatgat	cgtttttttag	gaatgcagcc	aaaaattgaa	ggttatgaca	gtagaaatac											837														
aagcaacagg	ctttgcttat	taaacatgta	attgaaaac													876														

<210> 40

<211> 230

<212> PRT

<213> Discosoma species

<400> 40

Met	Ser	Cys	Ser	Lys	Asn	Val	Ile	Lys	Glu	Phe	Met	Arg	Phe	Lys	Val
1				5					10					15	
Arg	Met	Glu	Gly	Thr	Val	Asn	Gly	His	Glu	Phe	Glu	Ile	Lys	Gly	Glu
			20					25					30		
Gly	Glu	Gly	Arg	Pro	Tyr	Glu	Gly	His	Cys	Ser	Val	Lys	Leu	Met	Val
		35					40					45			
Thr	Lys	Gly	Gly	Pro	Leu	Pro	Phe	Ala	Phe	Asp	Ile	Leu	Ser	Pro	Gln
	50					55					60				

-25-

Phe	Gln	Tyr	Gly	Ser	Lys	Val	Tyr	Val	Lys	His	Pro	Ala	Asp	Ile	Pro
65					70					75					80
Asp	Tyr	Lys	Lys	Leu	Ser	Phe	Pro	Glu	Gly	Phe	Lys	Trp	Glu	Arg	Val
				85					90					95	
Met	Asn	Phe	Glu	Asp	Gly	Gly	Val	Val	Thr	Val	Ser	Gln	Asp	Ser	Ser
			100					105					110		
Leu	Lys	Asp	Gly	Cys	Phe	Ile	Tyr	Glu	Val	Lys	Phe	Ile	Gly	Val	Asn
		115					120					125			
Phe	Pro	Ser	Asp	Gly	Pro	Val	Met	Gln	Arg	Arg	Thr	Arg	Gly	Trp	Glu
	130					135					140				
Ala	Ser	Ser	Glu	Arg	Leu	Tyr	Pro	Arg	Asp	Gly	Val	Leu	Lys	Gly	Asp
145					150					155					160
Ile	His	Met	Ala	Leu	Arg	Leu	Glu	Gly	Gly	Gly	His	Tyr	Leu	Val	Glu
				165				170						175	
Phe	Lys	Ser	Ile	Tyr	Met	Val	Lys	Lys	Pro	Ser	Val	Gln	Leu	Pro	Gly
			180					185					190		
Tyr	Tyr	Tyr	Val	Asp	Ser	Lys	Leu	Asp	Met	Thr	Ser	His	Asn	Glu	Asp
		195					200					205			
Tyr	Thr	Val	Val	Glu	Gln	Tyr	Glu	Lys	Thr	Gln	Gly	Arg	His	His	Pro
	210					215					220				
Phe	Ile	Lys	Pro	Leu	Gln										
225					230										

<210> 41
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> m-att;

<221> misc_difference
 <222> 18
 <223> n is a or g or c or t/u

<400> 41
 rkycwgcttt yktrtacnaa stsgb

25

<210> 42
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> m-attB;

<221> misc_difference
 <222> 18
 <223> n is a or g or c or t/u

<400> 42
 agccwgcttt yktrtacnaa ctsgb

25

<210> 43
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> m-attR

<221> misc_difference
 <222> 18
 <223> n is a or g or c or t/u

-26-

<400> 43
gttcagcttt cktrtacnaa ctsgb 25

<210> 44
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> m-attL

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

<400> 44
agccwgcttt cktrtacnaa gtsgb 25

<210> 45
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> m-attP1

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

<400> 45
gttcagcttt yktrtacnaa gtsgb 25

<210> 46
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB1

<400> 46
agcctgcttt tttgtacaaa cttgt 25

<210> 47
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB2

<400> 47
agcctgcttt cttgtacaaa cttgt 25

<210> 48
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB3

<400> 48
acccagcttt cttgtacaaa cttgt 25

<210> 49

-27-

<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR1

<400> 49
gttcagcttt tttgtacaaa cttgt 25

<210> 50
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR2

<400> 50
gttcagcttt cttgtacaaa cttgt 25

<210> 51
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR3

<400> 51
gttcagcttt cttgtacaaa gttgg 25

<210> 52
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL1

<400> 52
agcctgcttt tttgtacaaa gttgg 25

<210> 53
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL2

<400> 53
agcctgcttt cttgtacaaa gttgg 25

<210> 54
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL3

<400> 54
accagcttt cttgtacaaa gttgg 25

<210> 55
<211> 25

-28-

<212> DNA
 <213> Artificial Sequence

 <220>
 <223> attP1

 <400> 55
 gttcagcttt tttgtacaaa gttgg 25

 <210> 56
 <211> 25
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> attP2,P3

 <400> 56
 gttcagcttt cttgtacaaa gttgg 25

 <210> 57
 <211> 34
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Lox P site

 <400> 57
 ataacttcgt ataatgtatg ctatacgaag ttat 34

 <210> 58
 <211> 1032
 <212> DNA
 <213> Escherichia coli

 <220>
 <221> CDS
 <222> (1)...(1032)
 <223> nucleotide sequence encoding Cre recombinase

 <400> 58
 atg tcc aat tta ctg acc gta cac caa aat ttg cct gca tta ccg gtc 48
 Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val
 1 5 10 15

 gat gca acg agt gat gag gtt cgc aag aac ctg atg gac atg ttc agg 96
 Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg
 20 25 30

 gat cgc cag gcg ttt tct gag cat acc tgg aaa atg ctt ctg tcc gtt 144
 Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val
 35 40 45

 tgc cgg tcg tgg gcg gca tgg tgc aag ttg aat aac cgg aaa tgg ttt 192
 Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe
 50 55 60

 ccc gca gaa cct gaa gat gtt cgc gat tat ctt cta tat ctt cag gcg 240
 Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala
 65 70 75 80

 cgc ggt ctg gca gta aaa act atc cag caa cat ttg ggc cag cta aac 288
 Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn
 85 90 95

 atg ctt cat cgt cgg tcc ggg ctg cca cga cca agt gac agc aat gct 336

-29-

Met	Leu	His	Arg	Arg	Ser	Gly	Leu	Pro	Arg	Pro	Ser	Asp	Ser	Asn	Ala		
			100					105					110				
gtt	tca	ctg	gtt	atg	cgg	cgg	atc	cga	aaa	gaa	aac	gtt	gat	gcc	ggc		384
Val	Ser	Leu	Val	Met	Arg	Arg	Ile	Arg	Lys	Glu	Asn	Val	Asp	Ala	Gly		
		115					120					125					
gaa	cgt	gca	aaa	cag	gct	cta	gcg	ttc	gaa	cgc	act	gat	ttc	gac	cag		432
Glu	Arg	Ala	Lys	Gln	Ala	Leu	Ala	Phe	Glu	Arg	Thr	Asp	Phe	Asp	Gln		
	130					135					140						
gtt	cgt	tca	ctc	atg	gaa	aat	agc	gat	cgc	tgc	cag	gat	ata	cgt	aat		480
Val	Arg	Ser	Leu	Met	Glu	Asn	Ser	Asp	Arg	Cys	Gln	Asp	Ile	Arg	Asn		
145					150					155					160		
ctg	gca	ttt	ctg	ggg	att	gct	tat	aac	acc	ctg	tta	cgt	ata	gcc	gaa		528
Leu	Ala	Phe	Leu	Gly	Ile	Ala	Tyr	Asn	Thr	Leu	Leu	Arg	Ile	Ala	Glu		
			165					170						175			
att	gcc	agg	atc	agg	gtt	aaa	gat	atc	tca	cgt	act	gac	ggc	ggg	aga		576
Ile	Ala	Arg	Ile	Arg	Val	Lys	Asp	Ile	Ser	Arg	Thr	Asp	Gly	Gly	Arg		
			180					185					190				
atg	tta	atc	cat	att	ggc	aga	acg	aaa	acg	ctg	gtt	agc	acc	gca	ggc		624
Met	Leu	Ile	His	Ile	Gly	Arg	Thr	Lys	Thr	Leu	Val	Ser	Thr	Ala	Gly		
		195					200					205					
gta	gag	aag	gca	ctt	agc	ctg	ggg	gta	act	aaa	ctg	gtc	gag	cga	tgg		672
Val	Glu	Lys	Ala	Leu	Ser	Leu	Gly	Val	Thr	Lys	Leu	Val	Glu	Arg	Trp		
	210					215					220						
att	tcc	gtc	tct	ggc	gta	gct	gat	gat	ccg	aat	aac	tac	ctg	ttt	tgc		720
Ile	Ser	Val	Ser	Gly	Val	Ala	Asp	Asp	Pro	Asn	Asn	Tyr	Leu	Phe	Cys		
225					230					235					240		
cgg	gtc	aga	aaa	aat	ggc	gtt	gcc	gcg	cca	tct	gcc	acc	agc	cag	cta		768
Arg	Val	Arg	Lys	Asn	Gly	Val	Ala	Ala	Pro	Ser	Ala	Thr	Ser	Gln	Leu		
			245						250					255			
tca	act	cgc	gcc	ctg	gaa	ggg	att	ttt	gaa	gca	act	cat	cga	ttg	att		816
Ser	Thr	Arg	Ala	Leu	Glu	Gly	Ile	Phe	Glu	Ala	Thr	His	Arg	Leu	Ile		
			260					265					270				
tac	ggc	gct	aag	gat	gac	tct	ggc	cag	aga	tac	ctg	gcc	tgg	tct	gga		864
Tyr	Gly	Ala	Lys	Asp	Asp	Ser	Gly	Gln	Arg	Tyr	Leu	Ala	Trp	Ser	Gly		
		275					280					285					
cac	agt	gcc	cgt	gtc	gga	gcc	gcg	cga	gat	atg	gcc	cgc	gct	gga	gtt		912
His	Ser	Ala	Arg	Val	Gly	Ala	Ala	Arg	Asp	Met	Ala	Arg	Ala	Gly	Val		
	290					295					300						
tca	ata	ccg	gag	atc	atg	caa	gct	ggc	tgg	acc	aat	gta	aat	att			960
Ser	Ile	Pro	Glu	Ile	Met	Gln	Ala	Gly	Gly	Trp	Thr	Asn	Val	Asn	Ile		
305					310				315					320			
gtc	atg	aac	tat	atc	cgt	aac	ctg	gat	agt	gaa	aca	ggg	gca	atg	gtg		1008
Val	Met	Asn	Tyr	Ile	Arg	Asn	Leu	Asp	Ser	Glu	Thr	Gly	Ala	Met	Val		
				325					330				335				
cgc	ctg	ctg	gaa	gat	ggc	gat	tag										1032
Arg	Leu	Leu	Glu	Asp	Gly	Asp	*										
			340														

<210> 59
 <211> 343
 <212> PRT

-30-

<213> Escherichia coli

<400> 59

```

Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val
1          5          10          15
Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg
          20          25          30
Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val
          35          40          45
Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe
          50          55          60
Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala
          65          70          75          80
Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn
          85          90          95
Met Leu His Arg Arg Ser Gly Leu Pro Arg Pro Ser Asp Ser Asn Ala
          100          105          110
Val Ser Leu Val Met Arg Arg Ile Arg Lys Glu Asn Val Asp Ala Gly
          115          120          125
Glu Arg Ala Lys Gln Ala Leu Ala Phe Glu Arg Thr Asp Phe Asp Gln
          130          135          140
Val Arg Ser Leu Met Glu Asn Ser Asp Arg Cys Gln Asp Ile Arg Asn
          145          150          155          160
Leu Ala Phe Leu Gly Ile Ala Tyr Asn Thr Leu Leu Arg Ile Ala Glu
          165          170          175
Ile Ala Arg Ile Arg Val Lys Asp Ile Ser Arg Thr Asp Gly Gly Arg
          180          185          190
Met Leu Ile His Ile Gly Arg Thr Lys Thr Leu Val Ser Thr Ala Gly
          195          200          205
Val Glu Lys Ala Leu Ser Leu Gly Val Thr Lys Leu Val Glu Arg Trp
          210          215          220
Ile Ser Val Ser Gly Val Ala Asp Asp Pro Asn Asn Tyr Leu Phe Cys
          225          230          235          240
Arg Val Arg Lys Asn Gly Val Ala Ala Pro Ser Ala Thr Ser Gln Leu
          245          250          255
Ser Thr Arg Ala Leu Glu Gly Ile Phe Glu Ala Thr His Arg Leu Ile
          260          265          270
Tyr Gly Ala Lys Asp Asp Ser Gly Gln Arg Tyr Leu Ala Trp Ser Gly
          275          280          285
His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val
          290          295          300
Ser Ile Pro Glu Ile Met Gln Ala Gly Gly Trp Thr Asn Val Asn Ile
          305          310          315          320
Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Thr Gly Ala Met Val
          325          330          335
Arg Leu Leu Glu Asp Gly Asp
          340

```

<210> 60

<211> 1272

<212> DNA

<213> Saccharomyces cerevisiae

<220>

<221> CDS

<222> (1)...(1272)

<223> nucleotide sequence encoding Flip recombinase

<400> 60

```

atg cca caa ttt ggt ata tta tgt aaa aca cca cct aag gtg ctt gtt      48
Met Pro Gln Phe Gly Ile Leu Cys Lys Thr Pro Pro Lys Val Leu Val
1          5          10          15

cgt cag ttt gtg gaa agg ttt gaa aga cct tca ggt gag aaa ata gca      96
Arg Gln Phe Val Glu Arg Phe Glu Arg Pro Ser Gly Glu Lys Ile Ala
          20          25          30

```

-31-

tta	tgt	gct	gct	gaa	cta	acc	tat	tta	tgt	tgg	atg	att	aca	cat	aac	144
Leu	Cys	Ala	Ala	Glu	Leu	Thr	Tyr	Leu	Cys	Trp	Met	Ile	Thr	His	Asn	
		35					40					45				
gga	aca	gca	atc	aag	aga	gcc	aca	ttc	atg	agc	tat	aat	act	atc	ata	192
Gly	Thr	Ala	Ile	Lys	Arg	Ala	Thr	Phe	Met	Ser	Tyr	Asn	Thr	Ile	Ile	
	50					55				60						
agc	aat	tcg	ctg	agt	ttc	gat	att	gtc	aat	aaa	tca	ctc	cag	ttt	aaa	240
Ser	Asn	Ser	Leu	Ser	Phe	Asp	Ile	Val	Asn	Lys	Ser	Leu	Gln	Phe	Lys	
65					70					75					80	
tac	aag	acg	caa	aaa	gca	aca	att	ctg	gaa	gcc	tca	tta	aag	aaa	ttg	288
Tyr	Lys	Thr	Gln	Lys	Ala	Thr	Ile	Leu	Glu	Ala	Ser	Leu	Lys	Lys	Leu	
				85					90					95		
att	cct	gct	tgg	gaa	ttt	aca	att	att	cct	tac	tat	gga	caa	aaa	cat	336
Ile	Pro	Ala	Trp	Glu	Phe	Thr	Ile	Ile	Pro	Tyr	Tyr	Gly	Gln	Lys	His	
			100					105					110			
caa	tct	gat	atc	act	gat	att	gta	agt	agt	ttg	caa	tta	cag	ttc	gaa	384
Gln	Ser	Asp	Ile	Thr	Asp	Ile	Val	Ser	Ser	Leu	Gln	Leu	Gln	Phe	Glu	
		115					120					125				
tca	tcg	gaa	gaa	gca	gat	aag	gga	aat	agc	cac	agt	aaa	aaa	atg	ctt	432
Ser	Ser	Glu	Glu	Ala	Asp	Lys	Gly	Asn	Ser	His	Ser	Lys	Lys	Met	Leu	
		130				135					140					
aaa	gca	ctt	cta	agt	gag	ggg	gaa	agc	atc	tgg	gag	atc	act	gag	aaa	480
Lys	Ala	Leu	Leu	Ser	Gly	Glu	Glu	Ser	Ile	Trp	Glu	Ile	Thr	Glu	Lys	
145					150					155					160	
ata	cta	aat	tcg	ttt	gag	tat	act	tcg	aga	ttt	aca	aaa	aca	aaa	act	528
Ile	Leu	Asn	Ser	Phe	Glu	Tyr	Thr	Ser	Arg	Phe	Thr	Lys	Thr	Lys	Thr	
				165					170					175		
tta	tac	caa	ttc	ctc	ttc	cta	gct	act	ttc	atc	aat	tgt	gga	aga	ttc	576
Leu	Tyr	Gln	Phe	Leu	Phe	Leu	Ala	Thr	Phe	Ile	Asn	Cys	Gly	Arg	Phe	
			180					185					190			
agc	gat	att	aag	aac	ggt	gat	ccg	aaa	tca	ttt	aaa	tta	gtc	caa	aat	624
Ser	Asp	Ile	Lys	Asn	Val	Asp	Pro	Lys	Ser	Phe	Lys	Leu	Val	Gln	Asn	
		195					200					205				
aag	tat	ctg	gga	gta	ata	atc	cag	tgt	tta	gtg	aca	gag	aca	aag	aca	672
Lys	Tyr	Leu	Gly	Val	Ile	Ile	Gln	Cys	Leu	Val	Thr	Glu	Thr	Lys	Thr	
	210					215					220					
agc	gtt	agt	agg	cac	ata	tac	ttc	ttt	agc	gca	agg	ggg	agg	atc	gat	720
Ser	Val	Ser	Arg	His	Ile	Tyr	Phe	Phe	Ser	Ala	Arg	Gly	Arg	Ile	Asp	
225					230					235					240	
cca	ctt	gta	tat	ttg	gat	gaa	ttt	ttg	agg	aat	tct	gaa	cca	gtc	cta	768
Pro	Leu	Val	Tyr	Leu	Asp	Glu	Phe	Leu	Arg	Asn	Ser	Glu	Pro	Val	Leu	
				245					250					255		
aaa	cga	gta	aat	agg	acc	ggc	aat	tct	tca	agc	aat	aaa	cag	gaa	tac	816
Lys	Arg	Val	Asn	Arg	Thr	Gly	Asn	Ser	Ser	Ser	Asn	Lys	Gln	Glu	Tyr	
			260					265					270			
caa	tta	tta	aaa	gat	aac	tta	gtc	aga	tcg	tac	aat	aaa	gct	ttg	aag	864
Gln	Leu	Leu	Lys	Asp	Asn	Leu	Val	Arg	Ser	Tyr	Asn	Lys	Ala	Leu	Lys	
		275					280					285				
aaa	aat	gcg	cct	tat	tca	atc	ttt	gct	ata	aaa	aat	ggc	cca	aaa	tct	912
Lys	Asn	Ala	Pro	Tyr	Ser	Ile	Phe	Ala	Ile	Lys	Asn	Gly	Pro	Lys	Ser	
	290					295					300					

-32-

cac att gga aga cat ttg atg acc tca ttt ctt tca atg aag ggc cta	960
His Ile Gly Arg His Leu Met Thr Ser Phe Leu Ser Met Lys Gly Leu	
305 310 315 320	
acg gag ttg act aat gtt gtg gga aat tgg agc gat aag cgt gct tct	1008
Thr Glu Leu Thr Asn Val Val Gly Asn Trp Ser Asp Lys Arg Ala Ser	
325 330 335	
gcc gtg gcc agg aca acg tat act cat cag ata aca gca ata cct gat	1056
Ala Val Ala Arg Thr Thr Tyr Thr His Gln Ile Thr Ala Ile Pro Asp	
340 345 350	
cac tac ttc gca cta gtt tct cgg tac tat gca tat gat cca ata tca	1104
His Tyr Phe Ala Leu Val Ser Arg Tyr Tyr Ala Tyr Asp Pro Ile Ser	
355 360 365	
aag gaa atg ata gca ttg aag gat gag act aat cca att gag gag tgg	1152
Lys Glu Met Ile Ala Leu Lys Asp Glu Thr Asn Pro Ile Glu Glu Trp	
370 375 380	
cag cat ata gaa cag cta aag ggt agt gct gaa gga agc ata cga tac	1200
Gln His Ile Glu Gln Leu Lys Gly Ser Ala Glu Gly Ser Ile Arg Tyr	
385 390 395 400	
ccc gca tgg aat ggg ata ata tca cag gag gta cta gac tac ctt tca	1248
Pro Ala Trp Asn Gly Ile Ile Ser Gln Glu Val Leu Asp Tyr Leu Ser	
405 410 415	
tcc tac ata aat aga cgc ata taa	1272
Ser Tyr Ile Asn Arg Arg Ile *	
420	

<210> 61

<211> 422

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 61

Pro Gln Phe Gly Ile Leu Cys Lys Thr Pro Pro Lys Val Leu Val Arg	
1 5 10 15	
Gln Phe Val Glu Arg Phe Glu Arg Pro Ser Gly Glu Lys Ile Ala Leu	
20 25 30	
Cys Ala Ala Glu Leu Thr Tyr Leu Cys Trp Met Ile Thr His Asn Gly	
35 40 45	
Thr Ala Ile Lys Arg Ala Thr Phe Met Ser Tyr Asn Thr Ile Ile Ser	
50 55 60	
Asn Ser Leu Ser Phe Asp Ile Val Asn Lys Ser Leu Gln Phe Lys Tyr	
65 70 75 80	
Lys Thr Gln Lys Ala Thr Ile Leu Glu Ala Ser Leu Lys Lys Leu Ile	
85 90 95	
Pro Ala Trp Glu Phe Thr Ile Ile Pro Tyr Tyr Gly Gln Lys His Gln	
100 105 110	
Ser Asp Ile Thr Asp Ile Val Ser Ser Leu Gln Leu Gln Phe Glu Ser	
115 120 125	
Ser Glu Glu Ala Asp Lys Gly Asn Ser His Ser Lys Lys Met Leu Lys	
130 135 140	
Ala Leu Leu Ser Glu Gly Glu Ser Ile Trp Glu Ile Thr Glu Lys Ile	
145 150 155 160	
Leu Asn Ser Phe Glu Tyr Thr Ser Arg Phe Thr Lys Thr Lys Thr Leu	
165 170 175	
Tyr Gln Phe Leu Phe Leu Ala Thr Phe Ile Asn Cys Gly Arg Phe Ser	
180 185 190	
Asp Ile Lys Asn Val Asp Pro Lys Ser Phe Lys Leu Val Gln Asn Lys	
195 200 205	
Tyr Leu Gly Val Ile Ile Gln Cys Leu Val Thr Glu Thr Lys Thr Ser	

-33-

210	215	220
Val Ser Arg His Ile Tyr	Phe Phe Ser Ala Arg	Gly Arg Ile Asp Pro
225	230	235
Leu Val Tyr Leu Asp	Glu Phe Leu Arg Asn Ser	Glu Pro Val Leu Lys
245	250	255
Arg Val Asn Arg Thr	Gly Asn Ser Ser Ser	Asn Lys Gln Glu Tyr Gln
260	265	270
Leu Leu Lys Asp Asn	Leu Val Arg Ser Tyr	Asn Lys Ala Leu Lys Lys
275	280	285
Asn Ala Pro Tyr Ser	Ile Phe Ala Ile Lys	Asn Gly Pro Lys Ser His
290	295	300
Ile Gly Arg His Leu	Met Thr Ser Phe Leu	Ser Met Lys Gly Leu Thr
305	310	315
Glu Leu Thr Asn Val	Val Gly Asn Trp Ser	Asp Lys Arg Ala Ser Ala
325	330	335
Val Ala Arg Thr Thr	Tyr Thr His Gln Ile	Thr Ala Ile Pro Asp His
340	345	350
Tyr Phe Ala Leu Val	Ser Arg Tyr Tyr Ala	Tyr Asp Pro Ile Ser Lys
355	360	365
Glu Met Ile Ala Leu	Lys Asp Glu Thr Asn	Pro Ile Glu Glu Trp Gln
370	375	380
His Ile Glu Gln Leu	Lys Gly Ser Ala Glu	Gly Ser Ile Arg Tyr Pro
385	390	395
Ala Trp Asn Gly Ile	Ile Ser Gln Glu Val	Leu Asp Tyr Leu Ser Ser
405	410	415
Tyr Ile Asn Arg Arg	Ile	
420		

<210> 62
 <211> 48
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> IR2

<400> 62
 gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttc

48

<210> 63
 <211> 48
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> IR1

<400> 63
 gaagttccta tactttctag agaataggaa cttcgggaata ggaacttc

48

<210> 64
 <211> 66
 <212> DNA
 <213> Bacteriophage mu

<220>
 <221> CDS
 <222> (1)...(66)
 <223> nucleotide sequence encoding GIN recombinase

<400> 64	
tca act ctg tat aaa aaa cac ccc gcg aaa cga gcg cat ata gaa aac	48
Ser Thr Leu Tyr Lys Lys His Pro Ala Lys Arg Ala His Ile Glu Asn	
1 5 10 15	

gac gat cga atc aat taa
 Asp Asp Arg Ile Asn *

66

-34-

20

<210> 65
 <211> 21
 <212> PRT
 <213> bacteriophage mu

<400> 65
 Ser Thr Leu Tyr Lys Lys His Pro Ala Lys Arg Ala His Ile Glu Asn
 1 5 10 15
 Asp Asp Arg Ile Asn
 20

<210> 66
 <211> 69
 <212> DNA
 <213> Bacteriophage mu

<220>
 <221> CDS
 <222> (1)...(69)
 <223> nucleotide sequence encoding Gin recombinase

<400> 66
 tat aaa aaa cat ccc gcg aaa cga acg cat ata gaa aac gac gat cga 48
 Tyr Lys Lys His Pro Ala Lys Arg Thr His Ile Glu Asn Asp Asp Arg
 1 5 10 15
 atc aat caa atc gat cgg taa 69
 Ile Asn Gln Ile Asp Arg *
 20

<210> 67
 <211> 22
 <212> PRT
 <213> bacteriophage mu

<220>
 <223> Gin recombinase of bacteriophage mu

<400> 67
 Tyr Lys Lys His Pro Ala Lys Arg Thr His Ile Glu Asn Asp Asp Arg
 1 5 10 15
 Ile Asn Gln Ile Asp Arg
 20

<210> 68
 <211> 555
 <212> DNA
 <213> Escherichia coli

<220>
 <221> CDS
 <222> (1)...(555)
 <223> nucleotide sequence encoding PIN recombinase

<400> 68
 atg ctt att ggc tat gta cgc gta tca aca aat gac cag aac aca gat 48
 Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15
 cta caa cgt aat gcg ctg aac tgt gca gga tgc gag ctg att ttt gaa 96
 Leu Gln Arg Asn Ala Leu Asn Cys Ala Gly Cys Glu Leu Ile Phe Glu
 20 25 30

-35-

gac aag ata agc ggc aca aag tcc gaa agg ccg gga ctg aaa aaa ctg 144
 Asp Lys Ile Ser Gly Thr Lys Ser Glu Arg Pro Gly Leu Lys Lys Leu
 35 40 45
 ctc agg aca tta tcg gca ggt gac act ctg gtt gtc tgg aag ctg gat 192
 Leu Arg Thr Leu Ser Ala Gly Asp Thr Leu Val Val Trp Lys Leu Asp
 50 55 60
 cgg ctg ggg cgt agt atg cgg cat ctt gtc gtg ctg gtg gag gag ttg 240
 Arg Leu Gly Arg Ser Met Arg His Leu Val Val Leu Val Glu Glu Leu
 65 70 75 80
 cgc gaa cga ggc atc aac ttt cgt agt ctg acg gat tca att gat acc 288
 Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
 85 90 95
 agc aca cca atg gga cgc ttt ttc ttt cat gtg atg ggt gcc ctg gct 336
 Ser Thr Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala
 100 105 110
 gaa atg gag cgt gaa ctg att gtt gaa cga aca aaa gct gga ctg gaa 384
 Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Lys Ala Gly Leu Glu
 115 120 125
 act gct cgt gca cag gga cga att ggt gga cgt cgt ccc aaa ctt aca 432
 Thr Ala Arg Ala Gln Gly Arg Ile Gly Gly Arg Arg Pro Lys Leu Thr
 130 135 140
 cca gaa caa tgg gca caa gct gga cga tta att gca gca gga act cct 480
 Pro Glu Gln Trp Ala Gln Ala Gly Arg Leu Ile Ala Ala Gly Thr Pro
 145 150 155 160
 cgc cag aag gtg gcg att atc tat gat gtt ggt gtg tca act ttg tat 528
 Arg Gln Lys Val Ala Ile Ile Tyr Asp Val Gly Val Ser Thr Leu Tyr
 165 170 175
 aag agg ttt cct gca ggg gat aaa taa 555
 Lys Arg Phe Pro Ala Gly Asp Lys *
 180

<210> 69

<211> 184

<212> PRT

<213> Escherichia coli

<400> 69

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15
 Leu Gln Arg Asn Ala Leu Asn Cys Ala Gly Cys Glu Leu Ile Phe Glu
 20 25 30
 Asp Lys Ile Ser Gly Thr Lys Ser Glu Arg Pro Gly Leu Lys Lys Leu
 35 40 45
 Leu Arg Thr Leu Ser Ala Gly Asp Thr Leu Val Val Trp Lys Leu Asp
 50 55 60
 Arg Leu Gly Arg Ser Met Arg His Leu Val Val Leu Val Glu Glu Leu
 65 70 75 80
 Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
 85 90 95
 Ser Thr Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala
 100 105 110
 Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Lys Ala Gly Leu Glu
 115 120 125
 Thr Ala Arg Ala Gln Gly Arg Ile Gly Gly Arg Arg Pro Lys Leu Thr
 130 135 140
 Pro Glu Gln Trp Ala Gln Ala Gly Arg Leu Ile Ala Ala Gly Thr Pro
 145 150 155 160

-36-

Arg Gln Lys Val Ala Ile Ile Tyr Asp Val Gly Val Ser Thr Leu Tyr
 165 170 175
 Lys Arg Phe Pro Ala Gly Asp Lys
 180

<210> 70
 <211> 4778
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pcx plasmid

<400> 70
 gtcgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60
 gccatataat ggagttccgc gttacataac ttacggtaaa tggcccgccct ggctgaccgc 120
 ccaacgaccc cccgccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180
 ggactttcca ttgacgtcaa tgggtggact atttacggta aactgccacac ttggcagtac 240
 atcaagtgtat tcatatgcca agtacgcccc ctattgacgt caatgacggg aaatggcccg 300
 cctggcatta tgcccagtac atgaccttat gggactttcc tacttggcag tacatctacg 360
 tattagtcat cgcattacc atgggtcgag gtgagcccca cgttctgctt cactctcccc 420
 atctcccccc cctccccacc cccaattttg tattttatta ttttttaatt attttgtgca 480
 gcgatggggg cggggggggg ggggggcgcgc gccaggcggg gcggggcggg gcgagggggc 540
 gggcgggggc aggcgggagag gtgcggcggc agccaatcag agcggcgcgc tccgaaagtt 600
 tccttttatg gcgagggcgc ggcggcgggc gccctataaa aagcgaagcg cgcggcgggc 660
 gggagtgcgt gcgttgccct cgcgccgtgc cccgctccgc gccgcctcgc gccgcccgc 720
 ccggctctga ctgaccgcgt tactcccaca ggtgagcggg cgggacggcc cttctcctcc 780
 gggctgtaat tagcgcttg ttaaatgacg gctcgtttct tttctgtggc tgcgtgaaag 840
 ccttaaaggg ctccgggagg gccctttgtg cgggggggag cggctcgggg ggtgcgtgcg 900
 tgtgtgtgtg cgtggggagc gccgcgtgcg gccgcgcgtg cccggcggtg gtgagcgtg 960
 cgggcgcggc gcggggcctt gtgcgctccg cgtgtgcgcg aggggagcgc ggccgggggc 1020
 ggtgccccgc ggtgcggggg ggctgcgagg ggaacaaagg ctgcgtgcgg ggtgtgtgcg 1080
 tgggggggtg agcagggggg gtgggcgcgg cggtcgggct gtaaccccc cctgcacccc 1140
 cctccccgag ttgctgagca cggcccggtc tcgggtgcgg ggctccgtgc gggggcgtggc 1200
 gcgggcgcgc ccgtgcggg cgggggggtg cggcagtggt ggggtgccgg gggggcgggg 1260
 ccgcctcggg ccggggaggg ctccggggag gggcgcgcg catctccagc gcccgcggt 1320
 gtcgagggcg gcgcgacgc agccattgcc ttttatggta atcgtgcgag agggcgcgag 1380
 gacttccttt gtccaaaatc tggcggagcc gaaatctggg aggcgcgcgc gcacccccctc 1440
 tagcgggcgc gggcgaagcg gtgcggcgcc ggcaggaagg aaatgggcgg ggagggcctt 1500
 cgtgcgtcgc cgcgcgcgc tccccttctc catctccagc ctcggggctg ccgcaggggg 1560
 acggctgcct tcggggggga cggggcaggg cgggggttcg cttctggcgt gtgacggggc 1620
 gctctagagc ctctgctaac catgttcatg ccttcttctt tttcctacag ctctgggca 1680
 acgtgctggt tggtgtgctg tctcatcatt ttggcaaaaga attcactcct caggtgcagg 1740
 gtgcctatca gaaggtggtg gctgggtgtg ccaatgcctt ggctcacaaa taccactgag 1800
 atcttttttc ctctgccaaa aattatgggg acatcatgaa gcccttgag catctgactt 1860
 ctggctaata aaggaaattt attttcatgt caatagtgtg ttggaatttt ttgtgtctct 1920
 cactcggaaag gacatatggg agggcaaatc atttaaaaca tcagaatgag tatttggttt 1980
 agagtttggc aacatatgcc atatgctggc tgccatgaac aaaggtggct ataaagaggt 2040
 catcagtata tgaaacagcc cctgctgtc cattccttat tccatagaaa agccttgact 2100
 tgaggttaga ttttttttat attttgttt gtgttatttt tttctttaac atccctaaaa 2160
 ttttccttac atgttttact agccagattt ttctcctct cctgactact ccagtcata 2220
 gctgtccctc ttctcttatg aagatccctc gacctgcagc ccaagcttg cgtaatcatg 2280
 gtcatagctg tttcctgtgt gaaattgtta tccgctcaca attccacaca acatacagc 2340
 cggaagcata aagtgtaaag cctgggggtg ctaatgagtg agctaactca cattaattgc 2400
 gttgcgctca ctgcccgtt tccagtcggg aaacctgtcg tgccagcggg tccgcatctc 2460
 aattagtcag caaccatagt cccgccccta actccgccc tcccgcccct aactccgcc 2520
 agttccgccc attctccgc ccatggctga ctaatttttt ttatttatgc agaggccgag 2580
 gccgcctcgg cctctgagct attccagaag tagtgaggag gcttttttgg aggcctaggc 2640
 ttttgcaaaa agctaaactt ttatttgca cttataatgg ttacaaataa agcaatagca 2700
 tcacaaattt cacaaataaa gcattttttt cactgcattc tagttgtggt ttgtccaaac 2760
 tcatcaatgt atcttatcat gtctggatcc gctgcattaa tgaatcgcc aacgcgcggg 2820
 gagaggcggg ttgcgtattg ggcgctcttc cgcttccctc ctactgact cgctgcgtc 2880
 ggtcgttcgg gtccggcgag cggatcagc tcaactaaag gcggtaatac ggttatccac 2940
 agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggcagcaa 3000
 ccgtaaaaaa gccgcgttgc tggcggtttt ccataggctc cgcccccttg acgagcatca 3060
 caaaaatcga cgctcaagtc agaggtggcg aaaccgcaga ggactataaa gataccaggc 3120
 gtttccccct ggaagctccc tcgtgcgctc tcctgttccg accctgcgcg ttaccggata 3180

-37-

cctgtccgcc	tttctccctt	cgggaagcgt	ggcgctttct	caatgctcac	gctgtaggta	3240
tctcagttcg	gtgtaggtcg	ttcgctccaa	gctgggctgt	gtgcacgaac	cccccgttca	3300
gcccgaaccg	tgcgcccttat	ccggttaacta	tcgtcttgag	tccaacccgg	taagacacga	3360
cttatcgcca	ctggcagcag	ccactggtaa	caggattagc	agagcgagg	atgtaggcgg	3420
tgctacagag	ttcttgaagt	gggtggcctaa	ctacggctac	actagaagga	cagtatttgg	3480
tatctgcgct	ctgctgaagc	cagttacctt	cggaaaaaga	gttggtagct	cttgatccgg	3540
caaacaaacc	accgctggta	gcggtggttt	ttttgtttgc	aagcagcaga	ttacgcgcag	3600
aaaaaaagga	tctcaagaag	atcctttgat	cttttctacg	gggtctgacg	ctcagtgga	3660
cgaaaaactca	cgtaagggga	ttttgggtcat	gagattatca	aaaaggatct	tcacctagat	3720
ccttttaaat	taaaaatgaa	gttttaaatc	aatctaaagt	atatatgagt	aaacttggtc	3780
tgacagttac	caatgcttaa	tcagtgaggc	acctatctca	gcgatctgtc	tatttcgttc	3840
atccatagtt	gcctgactcc	ccgtcgtgta	gataactacg	atacgggagg	gcttaccatc	3900
tggccccagt	gctgcaatga	taccgcgaga	cccacgctca	ccggctccag	atztatcagc	3960
aataaaccag	ccagccggaa	gggcccagcg	cagaagtgg	cctgcaactt	tatccgcctc	4020
catccagtct	attaattggt	gccgggaagc	tagagtaagt	agttcgccag	ttaatatggtt	4080
gcgcaacgtt	gttgccattg	ctacaggcat	cgtggtgtca	cgctcgctcg	ttggtatggc	4140
ttcattcagc	tcgggttccc	aacgatcaag	gcgagttaca	tgatccccc	tggtgtgcaa	4200
aaaagcgggt	agctccttcg	gtcctccgat	cggtgtcaga	agtaagttgg	ccgcagtggt	4260
atcactcatg	gttatggcag	cactgcataa	ttctcttact	gtcatgccat	ccgtaagatg	4320
cttttctgtg	actgggtgag	actcaacca	gtcattctga	gaatagtgt	tgccggcacc	4380
gagttgctct	tgccggcggt	caataccgga	taataccgcg	ccacatagca	gaactttaaa	4440
agtgtctatc	attggaaaac	gttcttcggg	gcgaaaactc	tcaaggatct	taccgctggt	4500
gagatccagt	tcgatgtaac	ccactcgtgc	acccaactga	tcttcagcat	cttttacttt	4560
caccagcgtt	tctgggtgag	caaaaacagg	aaggcaaaat	gccgcaaaaa	agggaataag	4620
ggcgacacgg	aaatgttgaa	tactcatact	cttccttttt	caatattatt	gaagcattta	4680
tcagggttat	tgtctcatga	gcggatacat	atttgaatgt	atttagaaaa	ataaacaat	4740
aggggttccg	cgcacatttc	cccgaagaat	gccacctg			4778

<210> 71

<211> 5510

<212> DNA

<213> Artificial Sequence

<220>

<223> pCXeGFP plasmid

<400> 71

gtcgacattg	attattgact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gccatataat	ggagttccgc	gttacataac	ttacggtaaa	tggcccgcct	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtac	240
atcaagtgt	tcatatgcc	agtacgcccc	ctattgacgt	caatgacggt	aaatggccc	300
cctggcatta	tgcccagtac	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattagtc	tgctattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctcccc	420
atctcccccc	cctccccacc	cccaattttg	tatttattta	ttttttaatt	attttgtgca	480
gcgatggggg	cggggggggg	gggggcccgc	gccaggcggg	gcggggcggg	gcgagggggc	540
ggggcgggcg	aggcgagag	gtgcccgggc	agccaatcag	agcgggcggc	tccgaaagtt	600
tccttttatg	gcgaggcggc	ggcgggcggc	gccctataaa	aagcgaagcg	cgcgggcggc	660
gggagtcgct	gcgttgccct	cgccccgtgc	ccgctccgc	gccgcctcgc	gccgcccgc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtgagcggg	cgggacggcc	cttctcctcc	780
gggctgta	tagcgcttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgctgaaag	840
ccttaaagg	ctccgggagg	gccctttgtg	cgggggggag	cggtcggggg	gggtgcgtgc	900
tgtgtgtgtg	cgtggggagc	gccgcgtg	gcccgcgctg	cccggcggtc	gtgagcgtgc	960
cgggcgcggc	gcggggcttt	gtgcgtcccg	cgtgtgcgcg	aggggagcgc	ggcggggggc	1020
gggtgccccg	gggtgcgggg	ggctgcgagg	ggaacaaagg	ctgcgtgcgg	gggtgtgtgc	1080
tggggggggtg	agcagggggt	gtgggcgcgg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cggccccggt	tcgggtgcgg	ggctccgtgc	ggggcggtggc	1200
gcgggctcgc	ccgtgccggg	cgggggggtg	cggcgaggtg	gggtgcgggg	cggggcgggg	1260
ccgcctcggg	ccggggaggg	ctcgggggag	gggcgcggcg	gccccggagc	gccggcggtc	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttatggta	atcgtgcgag	agggcgcgag	1380
gacttccttt	gtcccaaact	tggcggagcc	gaaatctggg	aggcgccgc	gcacccccctc	1440
tagcggcggc	gggcgaagcg	gtgcggcgcc	ggcaggaagg	aaatgggcgg	ggaggggcctt	1500
cgtgcgtgcg	cgcgcgcgcg	tccccttctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acggctgcct	tcggggggga	cggggcaggg	cgggggttcg	cttctggcgt	gtgaccggcg	1620
gctctagagc	ctctgctaac	catgttcatt	ccttctctct	tttctacacg	ctcctgggca	1680
acgtgctgg	tgttgtgctg	tctcatcatt	ttggcaaaag	attcgccacc	atggtgagca	1740
agggcgagga	gctgttcacc	gggggtggtg	ccatcctggt	cgagctggac	ggcgacgtaa	1800

-38-

acggccacaa	gttcagcgtg	tccggcgagg	gcgagggcga	tgccacctac	ggcaagctga	1860
ccctgaagtt	catctgcacc	accggcaagc	tgcccgtgcc	ctggcccacc	ctcgtgacca	1920
ccttgacctt	cggtgtgcag	tgcttcagcc	gctaccccca	ccacatgaag	cagcacgact	1980
tcttcaagtc	cgccatgccc	gaaggctacg	tccaggagcg	caccatcttc	ttcaaggacg	2040
acggcaacta	caagaccggc	gocgaggtga	agttcgaggg	cgacaccctg	gtgaaccgca	2100
tcgagctgaa	gggcatcgac	ttcaaggagg	acggcaacat	cctggggcac	aagctggagt	2160
acaactacaa	cagccacaac	gtctatatca	tggccgacaa	gcagaagaac	ggcatcaagg	2220
tgaacttcaa	gatccgccac	aacatcgagg	acggcagcgt	gcagctcgcc	gaccactacc	2280
agcagaacac	ccccatcggc	gacggccccg	tgctgctgcc	cgacaaccac	tacctgagca	2340
cccagtcggc	cctgagcaaa	gaccccaacg	agaagcgcg	tcacatgggt	ctgctggagt	2400
tcgtagaccg	cgccgggatc	actctcgcca	tggacgagct	gtacaagtaa	gaattcactc	2460
ctcaggtgca	ggctgcctat	cagaagggtg	tggctggtgt	ggccaatgcc	ctggctcaca	2520
aataccactg	agatcttttt	ccctctgcca	aaaattatgg	ggacatcatg	aagccccttg	2580
agcatctgac	ttctggctaa	taaaggaaat	ttattttcat	tgcaatagtg	tgttggaatt	2640
ttttgtgtct	ctcactcgga	aggacatatg	ggagggcaaa	tcatttataa	catcagaatg	2700
agtattttgt	ttagagtttg	gcaacatatg	ccatatgctg	gctgccatga	acaaagggtg	2760
ctataaagag	gtcatcagta	tatgaaacag	cccctgctg	tccattcctt	attccataga	2820
aaagccttga	cttgagggtta	gatttttttt	atattttgtt	ttgtgttatt	tttttcttta	2880
acatccctaa	aattttcctt	acatgttttt	ctagccagat	ttttcctcct	ctcctgacta	2940
ctcccagtc	tagctgtccc	tcttctctta	tgaagatccc	tcgacctgca	gcccagctt	3000
ggcgtaatca	tggctatagc	tgtttcctgt	gtgaaattgt	tatccgctca	caattccaca	3060
caacatacga	gcccgaagca	taaagtgtaa	agcctggggg	gcctaataag	tgagctaact	3120
cacattaatt	gcgttgcgct	cactgcccgc	tttccagtcg	ggaaacctgt	cgtgccagcg	3180
gatccgcatc	tcaattagtc	agcaaccata	gtcccgcgcc	taactccgcc	catcccgcgc	3240
ctaactccgc	ccagttccgc	ccattctccg	ccccatggct	gactaatttt	ttttatttat	3300
gcagaggccg	agggccgcctc	ggcctctgag	ctattccaga	agtagtgagg	aggctttttt	3360
ggaggcctag	gcttttgcaa	aaagctaact	tgttttattg	agcttataat	ggttacaaat	3420
aaagcaatag	catcacaaat	ttcacaaata	aagcattttt	ttcactgcat	tctagtgtgt	3480
gtttgtccaa	actcatcaat	gtatcttatc	atgtctggat	ccgctgcatt	aatgaatcgg	3540
ccaacgcgcg	gggagaggcg	gtttgctgat	tgggcgctct	tccgcttcc	cgtcactga	3600
ctcgctgcgc	tccgttcgttc	ggctgcggcg	agcgttatca	gctcactcaa	aggcggtaat	3660
acggttatcc	acagaatcag	gggataacgc	aggaaagaac	atgtgagcaa	aaggccagca	3720
aaaggccagg	aaccgtaaaa	aggccgcggt	gctggcggtt	ttccataggc	tccgcccccc	3780
tgacgagcat	cacaaaaaat	gacgctcaag	tcagagggtg	cgaaacccga	caggactata	3840
aagataccag	gcgtttcccc	ctggaagctc	cctcgtgcgc	tctcctgttc	cgacctgccc	3900
gcttaccgga	tacctgtccg	cttttctccc	ttcgggaagc	gtggcgcttt	ctcaatgctc	3960
acgctgtagg	tatctcagtt	cgggtgtaggt	cgttcgctcc	aagctgggct	gtgtgcacga	4020
accccccggt	cagcccgacc	gctgcgcctt	atccggtaac	tatcgtcttg	agtccaaccc	4080
ggtaagacac	gacttatcgc	cactggcagc	agccactggg	aacaggatta	gcagagcgag	4140
gtatgtaggc	gggtctacag	agttcttgaa	gtgggtggct	aactacggct	acactagaag	4200
gacagtattt	ggtagctgag	ctctgctgaa	gccatttacc	ttcggaaaaa	gagttggtag	4260
ctcttgatcc	ggcaaacaaa	ccaccgctgg	tagcgggtgt	ttttttgttt	gcaagcagca	4320
gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	atctttttct	cggggctctga	4380
cgctcagtg	aacgaaaaat	cacgttaagg	gatttttggt	atgagattat	caaaaaggat	4440
cttcacctag	atcctttttaa	attaaaaatg	aagttttaaa	tcaatctaaa	gtatatatga	4500
gtaaaacttg	tctgacagtt	accaatgctt	aatcagtgag	gcacctatct	cagcgatctg	4560
tctatttcgt	tcatccatag	ttgcctgact	ccccgtcgtg	tagataacta	cgatacggga	4620
gggcttacca	tctggcccca	gtgctgcaat	gataccgcga	gacccacgct	caccggctcc	4680
agattttatc	gcaataaaac	agccagccgg	aagggccgag	gcgagaagtg	gtcctgcaac	4740
tttatccgcc	tccatccagt	ctattaattg	ttgcccggaa	gctagagtaa	gtagttcgcc	4800
agttaaatgt	ttgcgcaacg	ttgttgccat	tgctacaggc	atcgtgggtg	cacgctcgtc	4860
gttttggtatg	gcttcattca	gctccgggtc	ccaacgatca	aggcgagtta	catgatcccc	4920
catgttgtgc	aaaaaaagcg	ttagctcctt	cggctcctcg	atcgttgtca	gaagtaagtt	4980
ggccgcagtg	ttatcactca	tggttatggc	agcactgcat	aattctctta	ctgtcatgcc	5040
atccgtaaga	tgcttttctg	tgactgggtg	gtactcaacc	aagtcattct	gagaatagtg	5100
tatgcggcga	ccgagttgct	cttgcccggc	gtcaatacgg	gataataccg	cgccacatag	5160
cagaacttta	aaagtgtcoa	tcattggaaa	acgttcttcg	gggcgaaaac	tctcaaggat	5220
cttaccgctg	ttgagatcca	gttcgatgta	accactcgt	gcacccaact	gatcttcagc	5280
atcttttact	ttaccagcg	ttctgggtg	agcaaaaaa	ggaaggcaaa	atgcgcgaaa	5340
aaagggaata	agggcgacac	ggaaatgttg	aatactcata	ctcttctctt	ttcaatatta	5400
ttgaagcatt	tatcagggtt	attgtctcat	gagcggatac	atatttgaat	gtatttagaa	5460
aaataaacia	ataggggttc	cgcgcacatt	tcccgaaaaa	gtgccacctg		5510

<210> 72

<211> 282

<212> DNA

<213> Artificial Sequence

-39-

<220>

<223> attp

<400> 72

ccttgcgcta	atgctctgtt	acaggctcact	aataccatct	aagtagttga	ttcatagtga	60
ctgcataatgt	tgtgttttac	agtattatgt	agtctgtttt	ttatgcaaaa	totaatttaa	120
tatatggata	tttatatcat	tttacgtttc	tcgttcagct	tttttatact	aagttggcat	180
tataaaaaag	cattgcttat	caatttggtg	caacgaacag	gtcactatca	gtcaaaaataa	240
aatcattatt	tgatttcaat	tttgtccac	tcctgcctc	tg		282

<210> 73

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 73

ggccccgtaa	tcgagaagaa	20
------------	------------	----

<210> 74

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 74

ggtttaaagt	gcgctcctcc	aagaacgtca	tc	32
------------	------------	------------	----	----

<210> 75

<211> 40

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 75

agatctagag	cgcgcgtac	aggaacaggt	ggtggcggcc	40
------------	-----------	------------	------------	----

<210> 76

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer 5PacSV40

<400> 76

ctgttaatta	actgtggaat	gtgtgtcagt	taggggtg	37
------------	------------	------------	----------	----

<210> 77

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer Antisense Zeo

<400> 77

tgaacagggt	cacgtcgtcc	20
------------	------------	----

<210> 78

<211> 24

-40-

<212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer 5' HETS

<400> 78
 gggccgaaac gatctcaacc tatt 24

<210> 79
 <211> 19
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer 3' HETS

<400> 79
 cgcagcggcc ctcctactc 19

<210> 80
 <211> 29
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer 5BSD

<400> 80
 accatgaaaa catttaacat ttctcaaca 29

<210> 81
 <211> 29
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer SV40polyA

<400> 81
 tttatttggtg aaatttggtga tgctattgc 29

<210> 82
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer 3BSP

<400> 82
 ttaatttcgg gtatatttga gtgga 25

<210> 83
 <211> 32
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer EPO5XBA

<400> 83
 tatctagaat ggggggtgcac gaatgtcctg cc 32

<210> 84
 <211> 32
 <212> DNA

-41-

<213> Artificial Sequence

<220>

<223> Primer EPO3SBI

<400> 84

tacgtacgtc atctgtcccc tgtcctgcag gc

32

<210> 85

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer GENEPO3BSI

<400> 85

cgtacgtcat ctgtcccctg tcctgca

27

<210> 86

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer GENEPO5XBA

<400> 86

tctagaatgg ggggtgcacgg tgagtact

28

<210> 87

<211> 4862

<212> DNA

<213> Artificial Sequence

<220>

<223> pD2eGFP-1N plasmid from Clontech

<400> 87

tagttattaa	tagtaatcaa	ttacgggggc	attagttcat	agcccatata	tgaggttccg	60
cgttacataa	cttacggtaa	atggcccgcc	tggtgaccg	cccaacgacc	cccgccatt	120
gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	180
atgggtggag	tattttacgg	aaactgcccc	cttggcagta	catcaagtgt	atcatatgcc	240
aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgccagta	300
catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac	360
catggtgatg	cggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	420
atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	480
ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	gtaggcgtgt	540
acggtgggag	gtctatataa	gcagagctgg	tttagtgaac	cgtcagatcc	gctagcgcta	600
ccggactcag	atctcgagct	caagcttcga	attctgcagt	cgacggtacc	gcgggcccgg	660
gatccaccgg	tcgccaccat	ggtgagcaag	ggcgaggagc	tggtcaccgg	ggtggtgccc	720
atcctggtcg	agctggacgg	cgacgtaaac	ggccacaagt	tcagcgtgtc	cggcgagggg	780
gagggcgatg	ccacctacgg	caagctgacc	ctgaagttca	tctgcaccac	cggcaagctg	840
cccggtgccct	ggcccaccct	cgtgaccacc	ctgacctacg	gcgtgcagtg	cttcagccgc	900
taccccgacc	acatgaagca	gcacgacttc	ttcaagtccg	ccatgcccgga	aggctacgtc	960
caggagcgca	ccatcttctt	caaggacgac	ggcaactaca	agaccgcgc	cgaggtgaag	1020
ttcgagggcg	acaccctggt	gaaccgcata	gagctgaagg	gcatacgactt	caaggaggac	1080
ggcaacatcc	tggggcacia	gctggagtac	aactacaaca	gccacaacgt	ctatatcatg	1140
gccgacaagc	agaagaacgg	catcaagggtg	aacttcaaga	tcggccacaa	catcgaggac	1200
ggcagcgtgc	agctcgccga	ccactaccag	cagaacaccc	ccatcgccga	cggccccgtg	1260
ctgctgcccc	acaaccacta	cctgagcacc	cagtccgccc	tgagcaaaga	ccccaacgag	1320
aagcgcgatc	acatggctct	gctggagttc	gtgaccgccc	ccgggatcac	tctcggcatg	1380
gacgagctgt	acaagaagct	tagccatggc	ttcccgcgg	aggtggaggga	gcaggatgat	1440
ggcagcgtgc	ccatgtcttg	tgccaggag	agcgggatgg	accgtcacc	tgcagcctgt	1500
gcttctgcta	ggatcaatgt	gtagatgcgc	ggccgcgact	ctagatcata	atcagccata	1560
ccacatttgt	agaggtttta	cttgccttaa	aaaacctccc	acacctcccc	ctgaacctga	1620
aacataaaat	gaatgcaatt	gttgttgtta	acttgtttat	tgcagcttat	aatggttaca	1680

-42-

```

aataaagcaa tagcatcaca aatttcacaa ataaagcatt tttttcactg cattctagtt 1740
gtggtttgtc caaaactcatc aatgtatctt aaggcgtaaa ttgtaagcgt taatatatttg 1800
ttaaaattcg cggttaaattt ttgttaaattc agctcatttt ttaaccaata ggccgaaatc 1860
ggcaaaatcc cttataaatc aaaagaatag accgagatag ggttgagtggt tgttccagtt 1920
tggaacaaga gtccactatt aaagaacgtg gactccaacg tcaaagggcg aaaaaccgtc 1980
tatcagggcg atggcccact acgtgaacca tcaccctaata caagtttttt ggggtcgagg 2040
tgccgtaaaag cactaaatcg gaaccctaaa gggagccccc gatttagagc ttgacgggga 2100
aagccggcga acgtggcgag aaaggaaggg aagaaagcga aaggagcggg cgctagggcg 2160
ctggcaagtg tagcggtcac gctgcgcgta accaccacac ccgcccgcgt taatgcgcg 2220
ctacagggcg cgtcaggtgg cacttttcgg ggaaatgtgc gcggaacccc tatttgttta 2280
tttttctaaa tacattcaaa tatgtatccg ctcatgagac aataaccctg ataaatgctt 2340
caataatatt gaaaaaggaa gagtcctgag gcggaagaa ccagctgtgg aatgtgtgtc 2400
agttaggggt tggaaagtcc ccaggctccc cagcaggcag aagtatgcaa agcatgcac 2460
tcaatttagt agcaaccagg tgtggaaagt ccccaggctc ccagcaggc agaagtatgc 2520
aaagcatgca tctcaattag tcagcaacca tagtcccggc cctaactcgg cccatcccgc 2580
ccctaactcc gccagttcc gccatttctc cgccccatgg ctgactaatt ttttttattt 2640
atgcagaggc cgaggccgcc tggcctctg agctattcca gaagtagtga ggaggctttt 2700
ttggaggcct aggtttttgc aaagatcgat caagagacag gatgaggatc gtttcgcatg 2760
attgaacaag atggattgca cgcaggttct ccggccgctt ggggtggagag gctattcggc 2820
tatgactggg cacacagac aatcggctgc tctgtgccc cgtgttccg gctgtcagc 2880
caggggcgcc cggttctttt tgtcaagacc gtgcccgtga tgaactgcaa 2940
gacgaggcag cgcggctatc gtggctggcc acgacggcg ttccttgcc agctgtgtctc 3000
gacgttgtca ctgaagcggg aagggaactgg ctgctattgg gcgaagtgc ggggcaggat 3060
ctcctgtcat ctcaccttgc tcctgcccag aaagtatcca tcatggctga tgcaatgcgg 3120
ggctgcata cgcttgatcc ggctacctgc ccattcgacc accaagcgaa acatcgcatc 3180
gagcagcac gtactcggat ggaagccggg cttgtcgatc aggatgatct ggacgaagag 3240
catcaggggc tcgcgccagc cgaactgttc gccaggctca aggcgagcat gccgcagcgc 3300
gaggatctcg tcgtgaccca tggcgtatgc tgcttgccga atatcatggt ggaaaatggc 3360
cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg cggaccgcta tcaggacata 3420
gcgttggtca ccogtgatat tgctgaagag cttggcggcg aatgggctga ccgcttctc 3480
gtgctttacg gtatcgccgc tcccgaattc cagcgcatcg ccttctatcg ccttcttgac 3540
gagttcttct gagcgggact ctgggggttc aaatgaccga ccaagcgacg cccaacctgc 3600
catcacgaga ttctgatcc accgccgcct tctatgaaag gttgggcttc ggaatcgttt 3660
tccgggacgc cggctggatg atcctccagc gcggggatct catgctggag ttcttcgcc 3720
accctagggg gaggttaact gaaacacgga accggaagga acccgcgcta 3780
tgacggcaat aaaaagacag aataaaacgc acggtgttgg gtcgtttgtt cataaacgcg 3840
gggttcggtc ccagggtgg cactctgtcg atacccacc gagaccccat tggggccaat 3900
acgcccgcgt ttcttctctt tccccacccc acccccgaag ttccgggtgaa ggcccagggc 3960
tcgcagccaa cgtcggggcg gcaggccctg ccatagcctc aggttactca tatatactt 4020
agattgatit aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata 4080
atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag 4140
aaaagatcaa aggatcttct tgagatcctt ttttctgcg cgtaatctgc tgcttgcaa 4200
caaaaaaac accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactctt 4260
ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcct ctagtgtagc 4320
cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacatacctc gctctgctaa 4380
tcctgttacc agtggctgct gccagtggcg ataagtcgtg tcttacggg ttggactcaa 4440
gacgatagtt accggataag ggcagcggg cgggctgaac gggggggttcg tgcaacagc 4500
ccagcttggg gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa 4560
gcgccacgct toccgaagg agaaaggcgg acaggtatcc ggtaagcggc agggtcgga 4620
caggagagcg cagcagggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg 4680
ggtttcggca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc 4740
tatggaaaaa cgccagcaac gcggcctttt tacggttctt ggcccttttg tggccttttg 4800
ctcacatggt ctttctgcg ttatccctg attctgtgga taaccgtatt accgccatgc 4860
at

```

<210> 88

<211> 5192

<212> DNA

<213> Artificial Sequence

<220>

<223> pIRESpuro2 plasmid from Clontech

<400> 88

```

gacggatcgg gagatctccc gatcccctat ggtcgactct cagtacaate tgctctgatg 60
ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg 120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180

```

-43-

ttaggggttag	gcgtttttg	ctgcttcgcg	atgtacgggc	cagatatacg	cgttgacatt	240
gattattgac	tagttattaa	tagtaatacaa	ttacgggggtc	attagttcat	agcccatata	300
tggagttccg	cgttacataa	cttacggtaa	atggcccgcc	tggctgaccg	cccaacgacc	360
cccgccatt	gacgtcaata	atgacgtatg	ttcccatagt	aacgcccaata	gggactttcc	420
attgacgtca	atgggtggac	tatttacgggt	aaactgccca	cttggcgagta	catcaagtgt	480
atcatatgcc	aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	540
atgcccagta	catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	600
tcgctattac	catggtgatg	cggttttggc	agtacatcaa	tgggctgga	tagcggtttg	660
actcacgggg	atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	720
aaaatcaacg	ggactttcca	aaatgtcgt	acaactccgc	cccatgacg	caaagtggcg	780
gtaggcgtgt	acgggtgggag	gtctatataa	gcagagctct	ctggctaact	agagaacca	840
ctgcttactg	gcttatcgaa	attaatacga	ctcactatag	ggagacccaa	gcttgggtacc	900
gagctcggat	cgatatctgc	ggcctagcta	gcgcttaagg	cctgttaacc	ggctcgtacgt	960
ctccggattc	gaattcggat	ccgcggccgc	atagataact	gatccagtgt	gctggaatta	1020
attcgctgtc	tgcgagggcc	agctgttggg	gtgagtactc	cctctcaaaa	gcgggcatga	1080
cttctgcgct	aagattgtca	gtttccaaaa	acgaggagga	tttgatattc	acctggccc	1140
cggtgatgcc	tttgaggggtg	gccgcgtcca	tctggtcaga	aaagacaatc	ttttgttgt	1200
caagcttgag	gtgtggcagg	cttgagatct	ggccatacac	ttgagtgaca	atgacatcca	1260
ctttgccttt	ctctccacag	gtgtccactc	ccagggtccaa	ctgcaggtcg	agcatgcac	1320
tagggcgggc	aattccgccc	ctctccctcc	cccccccta	acgttactgg	ccgaagccgc	1380
ttggaataag	gcccgtgtgc	gtttgtctat	atgtgatttt	ccaccatatt	gccgtctttt	1440
ggcaatgtga	ggggccggaa	acctggccct	gtcttcttga	cgagcattcc	taggggtctt	1500
tccctctctg	ccaaaggaat	gcaaggctctg	ttgaatgtcg	tgaaggaagc	agttcctctg	1560
gaagcttctt	gaagacaaac	aacgtctgt	gcgacccttt	gcaggcagcg	gaacccccca	1620
cctggcgaca	ggtgcctctg	cggccaaaag	ccacgtgtat	aagatacacc	tgcaaaggcg	1680
gcacaacccc	agtgccacgt	tgtgagttgg	atagttgtgg	aaagagtcaa	atggctcttc	1740
tcaagcgtat	tcaacaaggg	gctgaaggat	gcccagaagg	tacccatttg	tatgggatct	1800
gatctggggc	ctcggtgcac	atgctttaca	tgtgtttagt	cgagggtaaa	aaaacgtcta	1860
ggccccccga	accacgggga	cgtggttttc	ctttgaaaaa	cacgatgata	agcttgccac	1920
aaccacaaag	gagacgacct	tccatgaccg	agtacaagcc	cacggtgcgc	ctcgccacc	1980
gcgacgcgt	cccccgggcc	gtacgcacc	tcgcgcgcgc	gttcgcccac	tacccgccca	2040
cgcgccacac	cgtcgaccgc	gaccgccaca	tcgagcgggt	caccgagctg	caagaactct	2100
tccctacgcg	cgtcgggctc	gacatcggca	aggtgtgggt	cgcggacgac	ggcgccgcgg	2160
tggcggtctg	gaccacgcgc	gagagcgtcg	aagcgggggc	ggtgttcgcc	gagatcggcg	2220
cgcgcattgc	cgagtttgagc	ggttcccgcc	tggccgcgca	gcaacagatg	gaaggcctcc	2280
tggcgccgca	ccggcccaag	gagcccgct	ggttccctggc	caccgtcggc	gtctcgcccg	2340
accaccaggg	caagggtctg	ggcagcgccg	tcgtgtctcc	cggagtggag	gcggccgagc	2400
gcgcgggggt	gcccgccttc	ctggagacct	ccgcgcgccg	caacctcccc	ttctacgagc	2460
ggctcggtct	caccgtcacc	gccgacgtcg	agtgcccgaa	ggaccgcgcg	acctgggtga	2520
tgaccgcgaa	gcccgggtgc	tgacgcgcgc	cccacgacc	gcagcgcccg	accgaaagga	2580
gcgcacgacc	ccatgggtcc	gaccgaagcc	gacccgggcg	gccccgcgca	ccccgcaccc	2640
gcccccgagg	cccaccgact	ctagagctcg	ctgatcagcc	tcgactgtgc	cttctagtgt	2700
ccagccatct	gttgtttgcc	cctccccctg	gccttccctg	acctggaag	gtgccactcc	2760
cactgtcctt	tcctaataaa	atgaggaat	tgcactcgcat	tgtctgagta	ggtgtcattc	2820
tattctgggg	ggtgggggtg	ggcaggacag	caagggggag	gattgggaag	acaattacag	2880
gcatgctggg	gatgcgggtg	gctctatggc	ttctgaggcg	gaaagaacca	gctggggctc	2940
gagtgcatcc	tagttgtggg	ttgtccaaac	tcatcaatgt	atcttatcat	gtctgtatac	3000
cgtcgacctc	tagctagagc	ttggcgtaat	catggtcata	gctgtttcct	gtgtgaaatt	3060
gttatccgct	cacaattcca	cacaacatac	gagccgggaag	cataaagtgt	aaagcctggg	3120
gtgcctaagt	agttagctaa	ctcacattaa	ttgcgttgcg	ctcactgccc	gctttccagt	3180
cgggaaacct	gtcgtgccag	ctgcattaat	gaatcgccca	acgcgcgggg	agaggcggtt	3240
tgcgtattgg	gcgctcttcc	gcttccctcg	tcactgactc	gctgcgctcg	gtcgttcggc	3300
tgcggcgagc	ggtatcagct	cactcaaagg	cggtataacg	gttatccaca	gaatcagggg	3360
ataacgcagg	aaagaacatg	tgagcaaaaag	gccagcaaaa	ggccagggaac	cgtaaaaagg	3420
ccgcgttgct	ggcggttttc	cataggctcc	gccccctga	cgagcatcac	aaaaatcgac	3480
gctcaagtca	gaggtggcga	aaccgcagac	gactataaag	ataccaggcg	tttccccctg	3540
gaagctccct	cgtgcgctct	cctgttccga	cctgcgcgct	taccggatac	ctgtccgcct	3600
ttctcccttc	gggaagcgtg	gcgctttctc	aatgctcacg	ctgtagggtat	ctcagttcgg	3660
tgtaggtcgt	tcgctccaag	ctgggctgtg	tgacgaaacc	ccccgttcag	cccagccgct	3720
gcgccttata	cggtaactat	cgtcttgagt	ccaaccgggt	aagacacgac	ttatcgccac	3780
tggcagcagc	cactggtaac	aggattagca	gagcgaggta	tgtaggcggt	gctacagagt	3840
tcttgaaagt	gtggcctaac	tacggctaca	ctagaaggac	agtatttggg	atctgcgctc	3900
tgctgaagcc	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	aaacaaacca	3960
ccgctgggtg	cgggtgggtt	tttgtttgca	agcagcagat	tacgcgcaga	aaaaaaggat	4020
ctcaagaaga	tcctttgatc	ttttctacgg	ggtctgacgc	tcagtggaac	gaaaactcac	4080
gttaagggat	tttgggtcatg	agattatcaa	aaaggatctt	cacctagatc	cttttaatt	4140
aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttgggtct	gacagttacc	4200

-44-

```

aatgcttaaat cagtgaggca cctatctcag cgaatctgtct atttcgttca tccatagttg 4260
cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct ggccccagtg 4320
ctgcaatgat accgcgagac ccacgctcac cgggtccaga tttatcagca ataaaccagc 4380
cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc atccagtcta 4440
ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg cgcaacgttg 4500
ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggatatggct tcattcagct 4560
ccggttcccc acgatcaagg cgagttacat gatcccccat gttgtgcaaa aaagcggtta 4620
gctccttcgg tctccgacac gttgtcagaa gtaagttggc cgcagtgtta tcactcatgg 4680
ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc ttttctgtga 4740
ctgggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg agttgctctt 4800
gcccggcgctc aatacgggat aataccgcgc cacatagcag aacttttaaa gtgctcatca 4860
ttggaaaacg tcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt 4920
cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc accagcgttt 4980
ctgggtgagc aaaaacagga aggcaaatg ccgcaaaaaa gggaataagg gcgacacgga 5040
aatgttgaat actcatactc ttcttttttc aatattattg aagcatttat cagggttatt 5100
gtctcatgag cggatacata tttgaatgta tttgaaaaa taaacaaata ggggttccgc 5160
gcacatttcc ccgaaaagtg ccacctgacg tc 5192

```

<210> 89

<211> 11182

<212> DNA

<213> Artificial Sequence

<220>

<223> pAg1 Plasmid

<400> 89

```

catgccaaacc acaggggttcc cctcgggac c aaagtacttt gatccaacc c tccgctgtct 60
atagtgcagt cggcttctga cgttcagtgc agcctgtctt tgaaaacgac atgtcgaca 120
agtcctaagt tacgcgacag gctgccgccc tgccttttct ctggcgtttt cttgtcgctg 180
gttttagctg cataaagtag aatacttgcc agatgaaccg gagacattac gccatgaaca 240
agagcgccgc cgttgccctg ctgggctatg ccgcgctcag caccgacgac caggacttga 300
ccaaccaacg ggccgaactg cacgcggccg gctgcacca gctgttttcc gagaagatca 360
ccggcaccag gcgcgaccgc ccggagctgg ccaggatgct tgaccaccta cgccctggcg 420
acgtttgtgac atgaccagg ctagaccgcc cggcccgag caccgcgac ctactggaca 480
ttgccgagcg catccaggag gccggcgccg gctgcgtag cctggcagag ccgtgggccc 540
acaccaccac gccggccggc cgcattggtg tgaccgtgtt cgccggcatt gccgagttcg 600
agcgttccct aatcatcgac cgcacccgga gcgggcgca ggccgccaag gcccgaggcg 660
tgaagtttgg ccccgccct accctcacc cggcacagat cgcgcacgcc cgcgagctga 720
tcgaccagga aggcgcgacc gtgaaagagg gcttgccact gcttgccgtg catcgctcga 780
ccctgtaccg cgcacttgag cgcagcgagg aagtgcgccc caccgaggcc aggcggcgcg 840
gtgccttccg tgaggacgca ttgaccgagg ccgacgccct ggccggccgc gagaatgaac 900
gccaagagga acaagcatga aaccgcacca ggacggccag gacgaaccgt ttttcattac 960
cgaagagatc gaggcggaga tgatcgcgcc cgggtacgtg ttcgagccgc ccgcgcagct 1020
ctcaaccgtg cggctgcatg aaatcctggc cgtttgtct gatgccaaagc tggcgccctg 1080
gccggccagc ttggccgctg aagaaaccga gcgcgcgct ctaaaaaggt gatgtgtatt 1140
tgagtaaaac agcttgctgc atgcggtcgc tgcgtatatg atgcgatgag taaataaaca 1200
aatacgcaag gggaacgcat gaaggttatc gctgtactta accagaaagg cgggtcaggc 1260
aagacacca tcgcaaccca tctagccgc gccctgcaac tcgcccgggc cgatgttctg 1320
ttagtgcatt ccgatcccca gggcagtgcc cgcgattggg cggcctgctg ggaagatcaa 1380
ccgctaaccg ttgtcgccat cgaccgcccg acgattgacc gcgacgtgaa ggccatcggc 1440
cggcgcgact tcgtagtgat cgacggagcg cccaggcgg cgacttgcc tgtgtccgcg 1500
atcaaggcag ccgacttcgt gctgattccg gtgcagccaa gcccttacga catatgggcc 1560
accgcccagc tgggtggagc ggttaagcag cgcattgagg tcacggatgg aaggctacaa 1620
gcccctttg tcgtgtcgcg ggcgatcaaa ggcacgcgca tcggcggtga ggttgccgag 1680
gcgctggccg ggtacgagc gccattctt gagtcccgt taacgcagcg cgtgagctac 1740
ccaggcactg ccgcgcgccc cacaaccgtt cttgaatcag aacccgaggg cgacgctgcc 1800
cgcgaggtcc aggcgctggc cgtgaaatt aaatcaaaac tcatttgagt taatgaggt 1860
aagagaaaat gagaaaacgc aaaaacacgc taagtccgg ccgtccgagc gcacccagca 1920
gcaaggctgc aacgttggcc agcctggcag acacgccagc catgaagcgg gtcaactttc 1980
agttgccggc ggaggatcac accaagctga agatgtacgc ggtacgcaa ggcaagacca 2040
ttaccgagct gctatctgaa tacatcgccg agtaccaga gtaaatgagc aaatgaataa 2100
atgagtgagc gaattttagc ggctaaaagg ggcggcatgg aaaatcaaga acaaccaggc 2160
accgacgccg tggaatgccc catgtgtgga ggaacgggcg gttggccagg cgtaagcgcc 2220
tggtgtgtct gccggccctg caatggcact ggaaccccca agcccagga atcgcgctga 2280
cggtcgcaaa ccatccggcc cgttacaaat cggcgccgcg ctgggtgatg acctggtgga 2340
gaagttgaag gccgcgcagg ccgccagcg gcaacgcacg gaggcagaag cacgccccg 2400

```


tgaatcgtgg	caagcggccg	ctgacgaat	ccgcaaagaa	tccccggcaac	cgccggcagc	2460
cgggtgcgccg	tcgattagga	agccgccccaa	gggcgacgag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tgcgacatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	ggtgatccgc	tacgagcttc	cagacgggca	2640
cgtagagggt	tccgcagggc	cggccggcat	ggccagtggt	tgggattacg	acctgggtact	2700
gatggcggtt	tcccatctaa	ccgaatccat	gaaccgatac	cgggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagtctctgcc	ggcgagccga	2820
tggcggaag	cagaaagacg	acctggtaga	aacctgcatt	cgttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggccgcctg	gtgacggtat	ccgaggggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaacc	ggcgggccgg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gacacagaa	ggcaagaacc	cggacgtgct	3060
gacggttcac	cccgattact	ttttgatcga	tcccggcatc	ggccgttttc	tctaccgcct	3120
ggcacgccgc	gccgcaggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gccggagagt	tcaagaagtt	ctgtttcacc	gtgcgcaagc	tgatcgggtc	3240
aatgacctg	ccggagtacg	atttgaagga	ggaggcgggg	caggctggcc	cgtacctagt	3300
catgcgctac	cgcgaacctga	tgcagggcga	agcatccgcc	ggttctaat	gtacggagca	3360
gatgctaggg	caaattgccc	tagcagggga	aaaaggctcg	aaaggctctc	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaaaccgt	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	catgtaagt	actgatataa	aagagaaaaa	3540
aggcgatttt	tccgcctaaa	actctttaaa	acttattaaa	actcttaaaa	cccgcctggc	3600
ctgtgcataa	ctgtctggcc	agcgcacagc	cgaagagctg	caaaaagcgc	ctacccttcg	3660
gtcgctgcgc	tccctacgcc	ccgcgccttc	gcgtcgccct	atcgcgcccg	ctggccgcctc	3720
aaaaatggct	ggcctacggc	caggcaatct	accaggcgcc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	cggcgccac	atcaaggcac	cctgcctcgc	gcgtttcggt	gatgacggtg	3840
aaaacctctg	acacatgcag	ctcccgga	cggtcacagc	ttgtctgtaa	gcggatgccg	3900
ggagcagaca	agcccgctcag	ggcgcgctcag	cgggtgttgg	cgggtgtcgg	ggcgagcca	3960
tgaccagtc	acgtagcgat	agcggagtgt	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcggtg	tgaaataccg	cacagatgcg	taaggagaaa	4080
ataccgcac	aggcgctctt	ccgcttctct	gctcactgac	tcgctgcgct	cggctcgctc	4140
gctgcggcga	gcggtatcag	ctcactcaaa	ggcggttaata	cggttatcca	cagaatcagg	4200
ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcggtg	ctggcggttt	tccataggct	ccgccccctt	gacgagcatc	acaaaaatcg	4320
acgctcaagt	cagagggtggc	gaaaccgcag	aggactataa	agataccagg	cgttttcccc	4380
tggaaagctc	ctcgtgcgct	ctcctgttcc	gacctgcgc	cttaccggat	acctgtctgc	4440
ctttctccct	tccgggaagcg	tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	4500
ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	ccccccggtt	agcccgaccg	4560
ctgcgcctta	tccggtaaact	atcgtcttga	gtccaaccgc	gtaaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tggtggccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4740
tctgctgaag	ccagttacct	tccgaaaaag	agttggtagc	tcttgatccg	gcaaacaac	4800
caccgctggt	agcggtggtt	tttttggttg	caagcagcag	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	gctcagtgga	acgaaaactc	4920
acgttaaggg	attttggta	tgcattctag	gtactaaaac	aattcatcca	gtaaaaata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	5040
ctgttcttcc	ccgatatact	ccctgatcga	cgggacgcag	aaggcaatgt	cataccactt	5100
gtccgccttg	ccgcttctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgctg	tctcccagggt	cgcggtggga	aaagacaagt	tctcttccgg	gcttttccgt	5220
ctttaaaaaa	tcatcacgct	cgccgggatc	tttaaatgga	gtgtcttctt	cccagttttc	5280
gcaatccaca	tccggccagat	cgttattcag	taagtaatcc	aattcggcta	agcggctgtc	5340
taagctattc	gtatagggac	aatccgatat	gtcgtaggag	tgaaagagcc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggtc	ttgttcactc	tcatactctt	ccgagcaaa	5460
gacgccatcg	gcctcactca	tgagcagatt	gctccagcca	tcatgccgtt	caaagtgcag	5520
gacctttgga	acagccagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5580
atcatagggtg	gtccctttat	accggctgtc	cgtcattttt	aaatataggt	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgcagcggta	5700
tttttcgatc	agttttttca	attccgggtga	tattctcatt	ttagccattt	attatttctc	5760
tctctttttc	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tcttgcattt	ctaaaacctt	aaacagcttt	aaacagcttt	ttcaagttg	5880
ttttcaaagt	tggcgataaa	catagtatcg	acggagccga	ttttgaaacc	ggcgtgatca	5940
caggcagcaa	cgctctgtca	tcggtacaat	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaacc	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaa	6060
tctgcccgcct	tacaacggct	ctcccgcgtga	cggcgcccg	gactgatggg	ctgctgtgat	6120
cgagtgggtga	ttttgtgccc	agctgcccgtg	cggggagctg	ttggctggct	gggtggcagga	6180
tatatgtgtg	tgtaaacaaa	ttgacgctta	gacaacttaa	taacacattg	cggacgtttt	6240
taatgtactg	aattaacgcc	gaattaatcc	gggggatctg	gatttttagta	ctggattttg	6300
gttttaggaa	ttagaaattt	tattgataga	agtattttac	aaatacaaat	acataactaag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctatagg	aaccctaatt	cccttatctg	6420

-46-

ggaactactc	acacattatt	atggagaaac	tcgagtcaaa	tctcgggtgac	gggcaggacc	6480
ggacggggcg	gtaccggcag	gctgaagtc	agctgccaga	aacccacgtc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgcctcgtgc	6600
atgcgcacgc	tccgggtcgtt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660
gcctccaggg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcccg	ccgctgggtg	6720
cgggggggga	cgtacacggg	cgactcggcc	gtccagtcgt	aggcgttgcg	tgccttccag	6780
gggccccgcg	aggcgatgcc	ggcgaacctg	ccgtccacct	cggcgacgag	ccagggatag	6840
cgtccccga	gacggacgag	gtcgtccgtc	cactcctgcg	gttcctgcgg	ctcgggtacgg	6900
aagttgaccg	tgtttgtctc	gatgtagtgg	ttgacgatgg	tgcagaccgc	cggcatgtcc	6960
gcctcgggtg	cacggcggat	gtcggccggg	cgtcgttctg	ggctcatggg	agactcgaga	7020
gagatagatt	tgtagagaga	gactggtgat	ttcagcgtgt	cctctccaaa	tgaatatgaac	7080
ttccttatat	agaggaagg	cttgcaagg	atagtggg	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaaagt	cttctttttc	7200
cacgatgctc	ctcgtgggtg	ggggtccatc	tttgggacca	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttcttttat	cgcaatgatg	gcattttag	gtgccacctt	ccttttctac	7320
tgtccttttg	atgaagtgc	agatagctgg	gcaattgga	ccgaggagg	ttcccgat	7380
taccctttgt	tgaaaagtct	caatagccct	ttgtcttct	gagactgtat	ctttgatatt	7440
cttgagtag	acgagagtgt	cgtgctccac	catgttatca	catcaatcca	cttgctttga	7500
agacgtgggt	ggaacgtctt	ctttttccac	gatgctcctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcttgaa	gatagccttt	cctttatcgc	aatgatggca	7620
tttgtagggt	ccaccttctt	tttctactgt	ccttttgatg	aagtgcaga	tagctgggca	7680
atggaaatccg	aggaggtttc	ccgatattac	ccttctgtga	aaagtctcaa	tagccctttg	7740
gtcttctgag	actgtatctt	tgatattctt	ggagtagacg	agagtgtcgt	gctccaccat	7800
gttggcaagc	tgtcttagcc	aatacgcaaa	ccgctctctc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcacgcacg	gtttcccgac	tggaaagcgg	gcagtgcg	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcacc	caggctttac	actttatgct	tccggctcgt	7980
atgttgtgtg	gaattgtgag	cggataacaa	tttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agccttgact	agagggtcga	cggatatacag	acatgataag	atacattgat	8100
gagtttggac	aaaccacaac	tagaatgcag	tgaaaaaaat	gctttatattg	tgaattttgt	8160
gatgctattg	ctttatattgt	aaccattata	agctgcaata	aacaagttgg	gggtgggcgaa	8220
gaactccagc	atgagatccc	cgcgctggag	gatcatccag	ccggcgctcc	ggaaaacgat	8280
tccgaagccc	aacctttcat	agaaggcggc	ggtggaatcg	aaatctcgta	gcacgtgtca	8340
gtcctgctcc	tcggccacga	agtgcacgca	gttgccggcc	gggtcgcgca	gggcgaactc	8400
ccgccccccac	ggctgctcgc	cgtctcgggt	catggccggc	ccggaggcgt	cccgggaagt	8460
cgtggacacg	acctccgacc	actcggcgta	cagctcgtcc	aggccgcgca	cccacacctc	8520
ggccagggtg	ttgtccggca	ccacctgggtc	ctggaccgcg	ctgatgaaca	gggtcacgtc	8580
gtcccggaac	acaccggcga	agtcgtcctc	cacgaagtcc	cgggagaacc	cgagccgggtc	8640
ggtccagaac	tcgaccgctc	cggcgacgtc	gcgcgcggtg	agcaccggaa	cggcactggg	8700
caacttggtc	atggatccag	atttcgctca	agttagtata	aaaaagcagg	cttcaatcct	8760
gcaggaattc	gatcgacact	ctcgtctact	caaagaatat	caaagataca	gtctcagaag	8820
accaaagggtc	tattgagact	tttcaacaaa	gggtaatatc	gggaaacctc	ctcggattcc	8880
attgccacgc	tatctgtcac	ttcatcaaaa	ggacagtaga	aaaggaagg	ggcacctaca	8940
aatgccatca	ttgcgataaa	ggaaaggcta	tcgttcaaga	tgcctctgcc	gacagtgggtc	9000
ccaaagattg	acccccaccc	acgaggagca	tcgtggaaaa	agaagacgtt	ccaaccacgt	9060
cttcaaaagca	agtggaattga	tgtgataaca	ttggtggagca	cgacactctc	gtctactcca	9120
agaatatcaa	agatacagtc	tcagaagacc	aaagggctat	tgagactttt	caacaaagg	9180
taatatcggt	aaacctcctc	ggattccatt	gcccagctat	ctgtcacttc	atcaaaagg	9240
cagtagaaaa	ggaagggtgg	acctacaaat	gccatcattg	cgataaagg	aaggctatcg	9300
ttcaagatgc	ctctgccgac	agtgggtccc	aagatggacc	cccacccacg	aggagcatcg	9360
tggaaaaaaga	agaogttcca	accacgtctt	caaagcaagt	ggattgatgt	gatattctcca	9420
ctgacgtaag	ggatgacgca	caatccact	atccttcgca	agaccttctt	ctatataagg	9480
aagttcattt	catttgagga	ggacacgctg	aaatcaccag	tctctctcta	caaattctatc	9540
tctctcgagc	tttcgcagat	ccgggggggg	aatgagatat	gaaaaagcct	gaactcaccg	9600
cgacgtctgt	cgagaagttt	ctgatcgaaa	agttcgacag	cgtctccgac	ctgatggcag	9660
tctcggaggg	cgaagaatct	cgtgctttca	gcttcgatgt	aggaggcgt	ggatatgtcc	9720
tgcgggtaaa	tagctgcgcc	gatggtttct	acaaagatcg	ttatgtttat	cggcactttg	9780
catcggccgc	gctcccgatt	ccggaagtgc	ttgacattgg	ggagtttagc	gagagcctga	9840
cctattgcat	ctcccgcgt	gcacagggtg	tcacgttgca	agacctgcct	gaaaccgaac	9900
tgcgcgctgt	tctacaaccg	ctcgcggagg	ctatggatgc	gatcgctgcg	gccgacttta	9960
gccagacgag	cgggttccgc	ccattcggac	cgcaagggaat	cgggtcaatac	actacatggc	10020
gtgatttcat	atgcgcgatt	gctgatcccc	atgtgtatca	ctggcaaac	gtgatggacg	10080
acaccgtcag	tgcgtccgtc	gcgacgggtc	tcgatgagct	gatgcttttg	gccgaggact	10140
gccccgaagt	ccggcacctc	gtgcacgcgg	atttccggctc	caacaatgtc	ctgacggaca	10200
atggccgcat	aacagcggtc	attgactgga	gcgagggcat	gttcggggat	tcccaatcag	10260
aggtcgccaa	catcttcttc	tggaggccgt	gggttgcttg	tatggagcag	cagacgcgct	10320
acttcgagcg	gaggcatccg	gagcttgcat	gatcgccacg	actccgggcg	tatatgctcc	10380
gcattgggtct	tgaccaactc	tatcagagct	tggttgacgg	caatttcgat	gatgcagctt	10440

-47-

```

gggcgcagggt  tccgatgcgac  gcaatcgtcc  gatccggagc  cgggactgtc  gggcgtagac  10500
aaatcgcccg  cagaagcgcg  gccgtctgga  ccgatggctg  tgtagaagta  ctcgccgata  10560
gtggaaacgg  acgccccagc  actcgtccga  gggcaaagaa  atagagtaga  tgccgaccgg  10620
atctgtcgat  cgacaagctc  gagtttctcc  ataataatgt  gtgagtagtt  cccagataag  10680
ggaattagggt  ttccctatagg  gtttcgctca  tgtgttgagc  atataagaaa  cccttagtat  10740
gtatttgat  ttgtaaaata  cttctatcaa  taaaatttct  aattcctaaa  accaaaatcc  10800
agtactaaaa  tccagatccc  ccgaattaat  tcggcggttaa  ttcagatcaa  gcttggcact  10860
ggccgctcgtt  ttacaacgct  gtgactggga  aaaccctggc  gttacccaac  ttaatcgcct  10920
tgcagcacat  ccccttttcg  ccagctggcg  taatagcgaa  gaggcccgca  ccgatcgccc  10980
ttcccaacag  ttgcgcgagg  tgaatggcga  atgctagagc  agcttgagct  tggatcagat  11040
tgtcgtttcc  cgccttcagt  ttaaactatc  agtgtttgac  aggatataat  ggcgggtaaa  11100
cctaagagaa  aagagcggtt  attagaataa  cggatattta  aaaggcggtg  aaaagggtta  11160
tccgctcgtc  catttgtagt  tg  11182

```

<210> 90

<211> 8428

<212> DNA

<213> Artificial Sequence

<220>

<223> pCambia3300 Plasmid

<400> 90

```

catgccaaac  acagggttcc  cctcgggata  aaagtacttt  gatccaaccc  ctccgctgct  60
atagtgcagt  cggcttctga  cgttcagtgc  agcgcgtctt  tgaaaacgac  atgtcgacac  120
agtcctaaat  tacgcgacag  gctgcgcgcc  tgcccttttc  ctggcggttt  cttgtcgcgt  180
gttttagtgc  cataaagtag  aatacttgcg  actagaaccg  gagacattac  gccatgaaca  240
agagcgccgc  cgctggcctg  ctgggctatg  cccgcgtcag  caccgacgac  caggacttga  300
ccaaccaacg  ggccgaactg  cacgcggccg  cctgcacca  gctgttttcc  gagaagatca  360
ccggcaccag  gcgcgacccg  ccggagctgg  ccaggatgct  tgaccaccta  cgccctggcg  420
acgttgtagc  agtgaccagg  ctgaccgcgc  tggcccgag  caccgcgcg  ctactggaca  480
ttgccgagcg  catccaggag  gccggcgccg  gcctgcgtag  cctggcagag  ccgtggggcg  540
acaccaccac  gccggccggc  cgcattggtg  tgaccgtgtt  cgccggcatt  gccgagttcg  600
agcgttccct  aatcatcgac  cgcacccgga  gcgggcgcga  ggccgcgaag  gcccgaggcg  660
tgaagtttgg  ccccgcctcc  accctcaccc  ccgcacagat  cgcgcacgcc  cgcgagctga  720
tcgaccagga  aggcgcgacc  gtgaaagagg  cggctgcact  gcttggcgtg  catcgctcga  780
ccctgtaccg  cgcacttgag  cgcagcgagg  aagtgcgccc  caccgaggcc  aggcggcgcg  840
gtgccttccg  tgaggacgca  ttgaccgagg  ccgacgccct  ggccggccgc  gagaatgaac  900
gccaagagga  acaagcatga  aaccgcacca  ggacggccag  gacgaaccgt  tttctattac  960
cgaagagatc  gaggcggaga  tgatcgccgc  cgggtacgtg  ttcgagccgc  ccgcgcacgt  1020
ctcaaccgtg  cggctgcatg  aaatcctggc  cgggttgtct  gatgccaagc  tggcgggcctg  1080
gccggccagc  ttggccgctg  aagaaaccga  gcgcgcgccg  ctaaaaagggt  gatgtgtatt  1140
tgagtaaaac  agcttgcgct  atgcggtcgc  tgcgtatatg  atgcgatgag  taaataaaca  1200
aatacgcaag  gggaaacgat  gaaggttatc  gctgtactta  accagaaagg  cgggtcaggc  1260
aagacgacca  tcgcaaccca  ctgagccgc  gccctgcaac  tcgcccgggg  ccatgtttctg  1320
ttagtcgatt  ccgatcccca  gggcagtgcc  cgcgattggg  cggcctgctg  ggaagatcaa  1380
ccgctaaccg  ttgtcggcat  cgaccgcccc  acgattgacc  gcgacgtgaa  ggccatcggc  1440
ggcgcgact  tcgtagtgat  cgacggagcg  cccagggcg  cggacttggc  tgtgtcccg  1500
atcaaggcag  cgcacttcgt  gctgattccg  gtgcagccaa  gcccttacga  catatgggcc  1560
accgcccacc  tgggtggagct  ggttaagcag  cgcattgagg  tcacggatgg  aaggctacaa  1620
gcggcctttg  tcgtgtcgcg  ggcgatcaaa  ggcacgcgca  tcggcggtga  ggttgccgag  1680
gcgctggccg  ggtacgagct  gccatttctt  gagtcccgt  tcacgcagcg  cgtgagctac  1740
ccaggcactg  ccgcccgcgg  cacaaccgtt  cttgaatcag  aaccgcaggg  cgacgctgcc  1800
cgcgaggtcc  aggcgctggc  cgctgaaatt  aaatcaaaac  tcatttgagt  taatgaggt  1860
aagagaaaat  gagcaaaagc  acaaacacgc  taagtgcggg  ccgtccgagc  gcacgcagca  1920
gcaaggctgc  aacgttggcc  agcctggcag  acacgccagc  catgaagcgg  gtcaactttc  1980
agttgcggcg  ggaggatcac  accaagctga  agatgtacgc  ggtacgccaa  ggcaagacca  2040
ttaccgagct  gctatctgaa  tacatcgccg  agctaccaga  gtaaatgagc  aaatgaataa  2100
atgagtagat  ggataaagga  ggctaaagga  ggcggcatgg  aaaatcaaga  acaaccaggc  2160
accgacggcg  tggaaatgcc  catgtgtgga  ggaacggcg  gttggccagg  cgtaagcggc  2220
tggtgtgtct  gccggccctg  caatggcact  ggaaccccca  agcccaggga  atcggcggtg  2280
cggctcgcaa  ccatccggcc  cggtacaaa  cggcgcgcg  ctgggtgatg  acctggtgga  2340
gaagttgaag  gccgcgcagg  ccgcccagcg  gcaaccgatc  gaggcagaag  cacgcccgg  2400
tgaatcgtgg  caagcgggcg  ctgatcgaat  ccgcaaagaa  tcccggcaac  cgccggcagc  2460
cgggtgcggc  tcgattagga  agccgcccga  gggcgacgag  caaccagatt  ttttcgttcc  2520
gatgctctat  gacgtgggca  cccgcgatag  tcgcagcatc  atggacgtgg  ccgttttccg  2580
tctgtcgaag  cgtgaccgac  gagctggcga  ggtgatccgc  tacgagcttc  cagacgggca  2640

```

cgtagaggtt	tccgcagggc	cggccggcat	ggccagtgtg	tgggattacg	acctgggtact	2700
gatggcgggt	tcccatctaa	ccgaatccat	gaaccgatac	cgggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaaag	cagaaagacg	acctggtaga	aacctgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggccgcctg	gtgacggtat	ccgaggggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaaac	ggcgggccgg	agtacatcga	3000
gacgagctga	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggacgtgct	3060
gacggttcac	cccgattact	ttttgatcga	tcccggcatc	ggccgttttc	tctaccgcct	3120
ggcacgcccgc	gcccgcaggca	aggcagaagc	cagatggttg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gccggagagt	tcaagaagtt	ctgtttcacc	gtgcgcaagc	tgatcgggtc	3240
aaatgacctg	ccggagtagc	atltgaagga	cgaggcgggg	caggctggcc	cgatcctagt	3300
catgcgctac	cgcaacctga	tgcagggcga	agcatccgcc	ggttccctaat	gtacggagca	3360
gatgctaggg	caaattgccc	tagcagggga	aaaaggctga	aaaggctctt	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaaaccgt	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	catgtaagt	actgatataa	aagagaaaaa	3540
aggcgatttt	tccgcctaaa	actctttaaa	acttattaaa	actcttaaaa	cccgcctggc	3600
ctgtgcataa	ctgtctggcc	agcgcacagc	cgaagagctg	caaaaagcgc	ctacccttcg	3660
gtcgtgcgcg	tccctacgcc	ccgccgcttc	gcgtcggcct	atcgcggccg	ctggccgctc	3720
aaaaatggct	ggcctacggc	caggcaatct	accagggcgc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	cggcgccac	atcaaggcac	cctgcctcgc	gcgtttcggg	gatgacgggt	3840
aaaacctctg	acacatgcag	ctcccggaga	cggtcacagc	ttgtctgtaa	gcgcatcgcc	3900
ggagcagaca	agcccgctcag	ggcgcgtcag	cgggtgttgg	cgggtgtcgg	ggcgcagcca	3960
tgacccagtc	acgtagegat	agcggagtg	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcggtg	tgaataccg	cacagatgcg	taaggagaaa	4080
ataccgatc	agggcgtctt	ccgcttcctc	gctcactgac	tcgctgcgct	cggtcgttcg	4140
gctcggcgga	gcggtatcag	ctcactcaaa	ggcggtaata	cggttatcca	cagaatcagg	4200
ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcgttg	ctggcggtttt	tccataggct	ccgccccctt	gacgagcatc	acaaaaatcg	4320
acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	agataaccagg	cgttttcccc	4380
tggaaagctc	ctcgtgcgct	ctcctgttcc	gacctgccc	cttaccggat	acctgtccgc	4440
ctttctccct	tccggaagcg	tggcgctttc	tcatagctca	cgctgtagg	atctcagttc	4500
ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	ccccccgttc	agcccgaccg	4560
ctgcgcctta	tccggtaact	atcgtcttga	gtccaaccgc	gtaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacgca	4680
gttctgaag	acttggccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4740
tctgctgaag	ccagttacct	tccgaaaaag	agttggtagc	tcttgatccg	gcaaaacaaac	4800
caccgctggg	agcgggtgggt	tttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gatcccttga	tcttttctac	ggggtctgac	gctcagtgga	acgaaaaactc	4920
acgttaaggg	atttttgcta	tgacttctag	gtactaaaac	aattcatcca	gtaaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	5040
ctgttcttcc	ccgatatact	ccctgatcga	ccggacgcag	aaggcaatgt	cataccactt	5100
gtccgcctcg	ccgcttctcc	caagatcaat	aaagccactt	actttgcca	ctttcacaaa	5160
gatgttgctg	tctcccagg	cgcctgtggg	aaagacaagt	tcctcttcgg	gcttttccgt	5220
cttaaaaaa	tcatacagct	cgcgcggatc	tttaaatgga	gtgtcttctt	cccggttttc	5280
gcaatccaca	tccggccagat	cgttattcag	taagtaatcc	aattcggcta	agcggctgtc	5340
taagctattc	gtatagggac	aatccgatat	gtcgatggag	tgaaagagcc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggct	ttgttcatct	tcatactctt	ccgagcaaac	5460
gacgccatcg	gcctcactca	tgagcagatt	gctccagcca	tcatgcggtt	caaagtcag	5520
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtctcttt	cccgttccac	5580
atcataggtg	gtccctttat	accggctgtc	cgtcattttt	aaatatagg	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgcagcggta	5700
tttttcgatc	agttttttca	attccgggtga	tattctcatt	ttagccattt	attatttctc	5760
tcctcttttc	tcacgtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tccttgcatt	ctaaaacctt	aaataccaga	aaacagcttt	ttcaaagttg	5880
ttttcaaagt	tggcgtataa	catagtatcg	acggagccga	ttttgaaacc	gcggtgatca	5940
caggcagcaa	cgctctgtca	tcgttacaa	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaacc	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaac	6060
tctgcgcgct	tacaacggct	ctcccgtga	cgccctcccg	gactgatggg	ctgcctgtat	6120
cgagtgtgga	ttttgtggcg	agctgccggg	cggggagctg	ttggctgggt	ggcggcagga	6180
tatatgtggt	tgtaaacaaa	ttgacgctta	gacaacttaa	taacacattg	cggacgtttt	6240
taatgtactg	aattaacgcc	gaattaatc	gggggatctg	gatttttagta	ctggattttg	6300
gttttaggaa	ttagaaattt	tattgataga	agtattttac	aaatacaaat	acatacttaag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctatagg	aaccttaatt	cccttatctg	6420
ggaactactc	acacattatt	atggagaaac	tcgagtcaaa	tctcggtgac	gggcaggacc	6480
ggacggggcg	gtaccggcag	gctgaagtcc	agctgccaga	aaccacagtc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgccctcgtg	6600
atgcgcacgc	tccgggtcgtt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660

-49-

```

gcctccaggg acttcagcag gtgggtgtag agcgtggagc ccagtcctgt ccgctgggtg 6720
cgggggggaga cgtacacggg cgactcggcc gtccagtcgt aggcgttgcg tgccttccag 6780
gggcccgcgt aggcgatgcc ggcgacctcg ccgtccacct cggcgacgag ccagggatag 6840
cgctcccgcg gacggacgag gtcgtccgtc cactcctgcg gttcctgctg ctcgggtacg 6900
aagttgaccg tgcttgtctc gatgtagtgg ttgacgatgg tgcagaccgc cggcatgtcc 6960
gcctcgggtg caccgaggat gtcggccggg cgtcgttctg ggctcatggg agactcgaga 7020
gagatagatt tgtagagaga gactggtgat ttcagcgtgt cctctccaaa tgaatgaac 7080
ttccttatat agaggaagg tttgcgaagg atagtgggat tgtgcgtcat cctttacgtc 7140
agtggagata tcacatcaat ccacttgctt tgaagacgtg gttggaacgt cttctttttc 7200
cacgatgctc ctgctgggtg ggggtccatc tttgggacca ctgtcggcag aggcattctt 7260
aacgatagcc tttcctttat cgcaatgatg gcatthgtag gtgccacctt cttttctac 7320
tgtccttttg atgaagtga agatagctgg gcaatggaat ccgaggagg tttccgatat 7380
taccctttgt tgaagaagt ctatagccct ttggtcttct gagactgtat ctttgatatt 7440
cttgagtag acgagagtgt cgtgctccac catgttatca catcaatcca cttgctttga 7500
agacgtggtt ggaacgtctt ctttttccac gatgctcctc gtgggtgggg gtccatcttt 7560
gggaccactg tcggcagagg catcttgaa cctttatcgc aatgatggca 7620
ttttaggtg ccaccttctt tttctactgt ccttttgatg aagtgcaga tagctgggca 7680
atggaatccg aggaggtttt ccgatattac cctttgttga aaagtctcaa tagccctttg 7740
gtcctctgag actgtatctt tgatattctt ggagtagacg agagtgtcgt gctccaccat 7800
gttggaagc tgctctagcc aatacgcaaa ccgcctctcc ccgcgcgtt gccgattcat 7860
taatgcagct ggcacgacag gtttcccgac tggaaagcgg gcagtgcgcg caacgcaatt 7920
aatgtgagtt agctcactca ttaggcaccc caggctttac actttatgct tccggctcgt 7980
atgttgtgtg gaattgtgag cggataacaa tttcacacag gaaacagcta tgaccatgat 8040
tacgaattcg agctcggtag ccggggatcc tctagagtcg acctgcaggc atgcaagctt 8100
ggcactggcc gtcgttttac aacgtcgtga ctgggaaaac cctggcgtaa cccaacttaa 8160
tcgccttgca gcacatcccc ctttcgccag ctggcgtaat agcgaagagg ccgcaccga 8220
tcgcctttcc caacagttgc gcagcctgaa tggcgaatgc tagagcagct tgagcttgga 8280
tcagattgtc gtttcccgc ttcagtttaa actatcagtg tttgacagga tatattggcg 8340
ggtaaaccta agagaaaaga gcgtttatta gaataacgga tatttaaaag ggcgtgaaaa 8400
ggtttatccg ttgcctcatt tgtatgtg 8428

```

<210> 91

<211> 3438

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attBZeo Plasmid

<400> 91

```

tcgaccctct agtcaaggcc ttaagtgagt cgtattacgg actggcgcgt gttttacaac 60
gtcgtgactg ggaaaacctt ggcgttaccc aacttaatcg ccttgcagca catccccctt 120
tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg ccttcccaa cagttgcgca 180
gctgaatgg cgaatggcgc ttcgcttggg aataaagccc gcttcggcgg gctttttttt 240
gttaactacg tcaggtggca cttttcgggg aaatgtgcgc ggaaccccta tttgtttatt 300
tttctaaata cattcaaata tgtatccgct catgagacaa taacctgat aaatgcttca 360
ataatattga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc ttattccctt 420
ttttgcggca ttttgccttc ctgtttttgc tcaccagaa acgctggtga aagtaaaaga 480
tgctgaagat cagttgggtg caccagtggg ttacatcgaa ctggatctca acagcggtaa 540
gatccttgag agttttcgcc ccgaagaacg ttctccaatg atgagcactt ttaaagtctt 600
gctatgtggc gcggtattat ccggtgttga cgccgggcaa gagcaactcg gtcgccgcat 660
acactattct cagaatgact tgggtgagta ctaccagtc acagaaaagc atcttacgga 720
tggcatgaca gtaagagaat tatgcagtgc tgccataacc atgagtgata acactgcgcg 780
caacttactt ctgacaacga tcggaggacc gaggagcta accgcttttt tgcacaacat 840
gggggatcat gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag ccataccaaa 900
cgacgagcgt gacaccacga tgcctgtagc aatggcaaca acgttgcgca aactattaac 960
tggcgaacta cttactctag cttcccggca acaattaata gactggatgg aggcggataa 1020
agttgcaggg ccactctctg gctcggccct tcggctggc tggttttatt ctgataaatt 1080
tggagccggg gagcgtgggt ctccgggtat cattgcagca ctggggccag atggtaagcc 1140
ctcccgatc gtagttatct acacgacggg gagtcaggca actatggatg aacgaaatag 1200
acagatcgct gagataggtg cctcactgat taagcattgg taactgtcag accaagttta 1260
ctcatatata ctttagattg atttaccctg gttgataatc agaaaagccc caaaaacagg 1320
aagattgtat aagcaaatat ttaaattgta aacgttaata ttttggttaa attcgcgtta 1380
aatttttgtt aaatcagctc attttttaac caataggccg aaatcggcaa aatcccttat 1440
aaatcaaaag aatagcccg gatagggttg agtggtgttc cagtttgga caagagtcca 1500
ctattaaaga acgtggactc caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc 1560
ccactacgtg aaccatcacc caaatcaagt tttttggggc cgaggtgccg taaagcacta 1620

```

-50-

```

aatcggaaacc ctaaagggag ccccccgattt agagcttgac ggggaaagcg aacgtggcga 1680
gaaaggaagg gaagaaaagcg aaaggagcgg gcgctagggc gctggcaagt gtagcgggtca 1740
cgctgcgcgt aaccaccaca ccgcgcgcgc ttaatgcgcc gctacagggc gcgtaaaagg 1800
atctaggtga agatcctttt tgataatctc atgaccaaaa tcccttaacg tgagttttcg 1860
ttccactgag cgtcagaccc cgtagaaaag atcaaaggat cttcttgaga tctttttttt 1920
ctgcgcgtaa tctgctgctt gcaaacaaaa aaaccaccgc taccagcggg ggtttgtttg 1980
ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata 2040
ccaaatactg ttcttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca 2100
ccgcctacat acctcgctct gctaatactg ttaccagtgg ctgctgccag tggcgataag 2160
tcgtgtctta ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc 2220
tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 2280
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aagggagaaa ggcggacagg 2340
tatccggtaa gcggcagggt cggaacagga gagcgcacga gggagcttcc agggggaac 2400
gcctgggtatc tttatagtcc tgtcgggttt cgcacactct gacttgagcg tcgatttttg 2460
tgatgctcgt cagggggggc gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg 2520
ttcctggcct tttgctggcc ttttgctcac atgtaatgtg agttagctca ctcataggc 2580
acccaggct ttacacttta tgcttccggc tcgtatgttg tgtggaattg tgagcggata 2640
acaatttcac acaggaaaca gctatgacca tgattacgcc aagctacgta atacgactca 2700
ctagtggggc ccgtgcaatt gaagccggct ggcgccaagc ttctctgcag gattgaagcc 2760
tgctttttta tactaacttg agcgaaatct ggatccatgg ccaagttgac cagtgccgtt 2820
ccggtgctca ccgcgcgcca cgtcgccgga gcggtcgagt tctggaccga ccggtcggg 2880
ttctcccggg acttcgtgga ggacgacttc gccggtgtgg tccgggacga cgtgaccctg 2940
ttcatcagcg cggccagga ccagggtggtg ccggacaaca ccctggcctg ggtgtgggtg 3000
cgcggcctgg acgagctgta cgccgagtg gtcggaggtcg tgtccacgaa cttccgggac 3060
gcctccgggc cggccatgac cgagatcggc gagcagccgt gggggcggga gttcccttg 3120
cgcgaccggc cgggcaactg cgtgcacttc gatggcagg agcaggactg acacgtgcta 3180
cgagatttcg attccaccgc cgccttctat gaaaggttg gcttcggaat cgttttccgg 3240
gacgcgggct ggtgatcct ccagcgcggg gatctcatgc tggagtctct cgcccacccc 3300
aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 3360
aataaagcat ttttttcaat gcattctagt tgtggtttgt ccaaactcat caatgtatct 3420
tatcatgtct gtataccg 3438

```

<210> 92

<211> 10549

<212> DNA

<213> Artificial Sequence

<220>

<223> pCambia1302 Plasmid

<300>

<308> Genbank #AF234398

<309> 2000-04-24

<400> 92

```

catggtagat ctgactagta aaggagaaga acttttctact ggagttgtcc caattcttgt 60
tgaattagat ggtgatgtta atggggcaca attttctgtc agtggagagg gtgaagggtga 120
tgcaacatac ggaaaactta cccttaaatt tatttgcact actggaaaac tacctgttcc 180
gtggccaaca cttgtcacta ctttctotta tgggtgttcaa tgcttttcaa gataccaga 240
tcatatgaag cggcacgact tcttcaagag cgccatgcct gagggatacg tgcaggagag 300
gaccatcttc ttcaaggacg acgggaacta caagacacgt gctgaagtca agtttgaggg 360
agacaccctc gtcaacagga tcgagcttaa gggaatcgat ttcaaggagg acggaaacat 420
cctcggccac aagttggaat acaactacaa ctcccacac gtatacatca tggccgacaa 480
gcaaaaagaac ggcatacaaa ccaacttcaa gaccgcgac aacatcgaag acggcgggct 540
gcaactcgct gatcattatc aacaaaatac tccaattggc gatggccctg tctttttacc 600
agacaaccat tacctgtcca cacaatctgc cctttcgaaa gatcccaacg aaaagagaga 660
ccacatgggt cttcttgagt ttgtaacagc tgctgggatt acacatggca tggatgaact 720
atacaaaagt agccaccacc accaactacc cgttggaatt ggtgaccagc tcgaatttcc 780
ccgatcggtc aaacatttgg caataaagtt cttaagatt gaatcctgtt gccggtcttg 840
cgatgattat catataattt ctgttggaatt acgttaagca tgtaataatt aacatgtaat 900
gcatgacgtt atttatgaga tgggttttta tgattagagt cccgcaatta tacatttaat 960
acgcgataga aaacaaaata tagcgcgcaa actaggataa attatcgcgc gcggtgtcat 1020
ctatgttact atagtcgggaa ttaaaactatc agtgtttgac aggatataat ggcgggtaaa 1080
cctaagagaa aagagcggtt attagaataa cggatatatta aaagggcggtg aaaaggttta 1140
tccgttcgtc catttgatg tgcatgccaa ccacagggtt cccctcggga tcaaagtact 1200
ttgatccaac ccctccgctg ctatagtga gtcggcttct gacgttcagt gcagccgtct 1260
tctgaaaacg acatgtcgca caagtcctaa gttacgcgac aggttgccgc cctgcccttt 1320

```

-51-

tccctggcggtt	ttcttgtgcg	gtgttttagt	cgcataaagt	agaatacttg	cgactagaac	1380
cggagacatt	acgccatgaa	caagagcgcc	gccgctggcc	tgctgggcta	tgcccgcgtc	1440
agcaccgacg	accaggactt	gaccaaccac	cgggcccgaac	tgcaacgcggc	cggctgcacc	1500
aagctgtttt	ccgagaagat	caccggcacc	aggcgcgacc	gcccggagct	ggccaggatg	1560
cttgaccacc	tacgccttgg	cgacgttgtg	acagtgcacca	ggctagaccg	cctggccccg	1620
agcaccgcg	acctactgga	cattgcccag	cgcatccagg	aggccggcgc	gggcctgctg	1680
agcctggcag	agccgtgggc	cgacaccacc	acgccggccg	gccgcatggg	gttgaccctg	1740
ttcgccggca	ttgcgaggtt	cgagcgttcc	ctaatacatg	accgcacccg	gagcggggcg	1800
gaggccgcca	aggcccgagg	cgtgaagttt	ggcccccgcc	ctaccctcac	cccggcacag	1860
atcgcgacag	cccgcgagct	gatcgaccag	gaaggccgca	ccgtgaaaga	ggcggctgca	1920
ctgcttggcg	tgcatcgctc	gacctgttac	cgcgcacttg	agcgcagcga	ggaagtgcag	1980
cccaccgagg	ccaggcggcg	cgggtgccttc	cgtgaggacg	cattgaccga	ggccgacgcc	2040
ctggcggccg	ccgagaatga	acgccaagag	gaacaagcat	gaaaccgcac	caggacggcc	2100
aggacgaacc	gtttttcatt	accgaagaga	tcgaggcggg	gatgatcgcg	gccgggtacg	2160
tgcttcgagcc	gcccgcgcac	gtctcaaccg	tgccgctgca	tgaatatcctg	gccgggttgt	2220
ctgatgccaa	gctggcggcc	tggccggcca	gcttggccgc	tgaagaaacc	gagcgcgcgc	2280
gtctaaaaag	gtgatgtgta	tttgagttaa	acagtgtcgc	tcattgcggtc	gctgcgtata	2340
tgatgcatg	agtaataaaa	caaatacgca	aggggaacgc	atgaaggtta	tcgctgtact	2400
taaccagaaa	ggcgggtcag	gcaagacgac	catcgcaacc	catctagccc	gcgccttgca	2460
actcgccggg	gccgatgttc	tgttagtcca	ttccgatccc	cagggcagtg	cccgcgattg	2520
ggcggccgctg	cgggaagatc	aaccgcatac	cgttgtcggc	atcgaccgcc	cgacgattga	2580
ccgcgacgtg	aaggccatcg	gccggcgcga	cttcgtagt	atcgacggag	cgcgccaggc	2640
ggcggacttg	gctgtgtccg	cgatcaaggc	agccgacttc	gtgctgattc	cgggtgcagcc	2700
aagcccttac	gacatatggg	ccaccgccga	cctggtggag	ctgggttaagc	agcgcattga	2760
ggtcacggat	ggaaggctac	aagcggcctt	tgtcgtgtcg	cggggcatca	aaggcacgcg	2820
catcgccggt	gaggttgccg	aggcgtggc	cgggtacgag	ctgcccattc	ttgagtcctg	2880
tatcacgcag	cgcgtgagct	accaggcac	tgccgccgcc	ggcacaaccg	ttcttgaatc	2940
agaaccgag	ggcgacgctg	cccgcgaggt	ccaggcgtg	gccgctgaaa	ttaaatcaaa	3000
actcatttga	gttaatgagg	taaagagaaa	atgagcaaaa	gcacaaacac	gctaagtgcc	3060
ggccgtccga	gcgcacgcag	cagcaaggct	gcaacgttgg	ccagcctggc	agacacgcga	3120
gccatgaagc	gggtcaactt	tcagttgcgc	cggaggatc	acaccaagct	gaagatgtac	3180
gcgttacgcc	aaggcaagac	cattaccgag	ctgctatctg	aatacatcgc	gcagctacca	3240
gagtaaatga	gcaaatgaat	aaatgagtag	atgaatttta	gcggctaaag	gaggcggcat	3300
ggaaaatcaa	gaacaaccag	gcaccgacgc	cgtggaatgc	cccatgtgtg	gaggaacggg	3360
cggttggcca	ggcgtaagcg	gctgggttgt	ctgccggccc	tgcaatggca	ctggaacccc	3420
caagcccgag	gaatcggcgt	gacggtcgca	aaccatccgg	cccggtacaa	atcggcgccg	3480
cgctgggtga	tgacctgggtg	gagaagttga	aggccgcgca	ggccgcccag	cggcaacgca	3540
tcgaggcaga	agcacgcccc	ggtgaatcgt	ggcaagcgcc	cgctgatcga	atccgcaaa	3600
aatcccggca	accgcgggca	gccgggtgcg	cgtcgattag	gaagccgccc	aaggggcgacg	3660
agcaaccaga	ttttttcggt	ccgatgcttc	atgacgtggg	caccocgcat	agtcgcagca	3720
tcattggacgt	ggccggtttt	cgtctgtcga	agcgtgaccg	acgagctggc	gaggtgatcc	3780
gctacgagct	tcagacgggg	cacgtagagg	tttccgcagg	gccggccggc	atggccagtg	3840
tgtgggatta	cgacctgggt	ctgatggcgg	tttcccatct	aaccgaatcc	atgaaccgat	3900
accgggaagg	gaagggagac	aagccccggc	gcgtgttccg	tcacacagtt	gccgacgtac	3960
tcaagtctcg	ccggcgagcc	gatggcggaa	ggcaaaaaga	cgacctggta	gaaacctgca	4020
ttcggttaaa	caccacgcac	ggtgccatgc	agcgtacgaa	gaaggccaag	aacggccgcc	4080
tggtgacggt	atccgagggg	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	4140
ccgggcggcc	ggagtacatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	4200
aaggcaagaa	cccggacgtg	ctgacggttc	accccgatta	ctttttgatc	gatcccggca	4260
tcggccggtt	tcctctacgc	ctggcacgcc	cgccgcgagg	caaggcagaa	gccagatggg	4320
tgttcaagac	gatctacgaa	cgcagtggca	gcgccggaga	gttcaagaag	ttctgtttca	4380
ccgtgcgcaa	gctgatcggg	tcaaatagac	tgccggagta	cgatttgaag	gaggaggcgg	4440
ggcaggctgg	cccgatccct	gtcatgcgct	accgcaacct	gatcgagggc	gaagcatccg	4500
ccggttccta	atgtacggag	cagatgctag	ggcaaatgct	cctagcaggg	gaaaaagggtc	4560
gaaaaggctc	ctttcctgtg	gatagcacgt	acattgggaa	cccaaaggccg	tacattggga	4620
accggaaccc	gtacattggg	aacccaaagc	cgtacattgg	gaaccgggtca	cacatgtaag	4680
tgactgatat	aaaagagaaa	aaaggcgatt	tttccgccta	aaactottta	aaacttatta	4740
aaactcttaa	aaaccgcctg	gcctgtgcac	aactgtctgg	ccagcgcaca	gccgaagagc	4800
tgcaaaaagc	gcctaccctt	cggtcgctgc	gctccctacg	ccccgcgct	tcgcgtcgcc	4860
ctatcgccgg	cgttggccgc	tcaaaaaatg	ctggcctacg	gccaggcaat	ctaccagggc	4920
gcggacaagc	cgcgcgcgtc	ccactcgacc	gccggcgccc	acatcaaggc	accctgcctc	4980
gcgcgtttcg	gtgatgacgg	tgaaaacctc	tgacacatgc	agctcccggg	gacgggtcac	5040
gcttggctgt	aagcggatgc	cgggagcaga	caagcccgct	agggcgcgct	agcgggtggt	5100
ggcgggtgtc	ggggcgcgagc	catgacccag	ctagcgtagc	atagcggagt	gtatactggc	5160
ttaactatgc	ggcatcagag	cagattgtac	tgagagtgc	ccatatgcgg	tgtgaaatac	5220
cgcacagatg	cgtaaggaga	aaataccgca	tcaggcgtc	ttccgcttcc	tcgctcactg	5280
actcgtcgcg	ctcggtcggt	cggctgcggc	gagcgggtatc	agctcactca	aaggcggtaa	5340

-52-

tacggttatc	cacagaatca	ggggataacg	caggaaagaa	catgtgagca	aaaggccagc	5400
aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggt	tttccatagg	ctccgcccc	5460
ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagagggtg	gcgaaaccg	acaggactat	5520
aaagatacca	ggcgtttccc	cctggaagct	ccctcgtgcg	ctctcctggt	ccgaccctgc	5580
cgcttacccg	atacctgtcc	gcctttctcc	cttcgggaag	cgtggcgctt	tctcatagct	5640
cacgctgtag	gtatctcagt	tcgggtgtagg	tcggtcgtcc	caagctgggc	tgtgtgcacg	5700
aacccccgct	tcagcccgac	cgctgcgcct	tatccggtaa	ctatcgtctt	gagtccaacc	5760
cggtaaagca	cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	5820
ggatgtagg	cggtgctaca	gagttcttga	agtgggtggc	taactacggc	tacactagaa	5880
ggacagtatt	tggtatctgc	gctctgctga	agccagttac	cttcggaaaa	agagttggta	5940
gctcttgatc	cggaacaaa	accaccgctg	gtagcgggtg	tttttttgtt	tgcaagcagc	6000
agattacgcg	atctttcaca	ggatctcaag	aagatccttt	gatcttttct	acggggtctg	6060
acgctcagtg	gaacgaaaac	tcacgttaag	ggatttttgt	catgcattct	aggtactaaa	6120
acaattcatc	cagtaaaata	taatatttta	ttttctccca	atcaggcttg	atccccagta	6180
agtcacaaaa	tagctcgaca	tactgttctt	ccccgatata	ctccctgata	gaccggacgc	6240
agaaggcaat	gtcataccac	ttgtccgcct	tgccgcttct	cccaagatca	ataaagccac	6300
ttactttgoc	atctttcaca	aagatgttgc	tgtctcccag	gtcgcggtgg	gaaaagacaa	6360
gttccctctt	gggcttttcc	gtctttaaaa	aatcatacag	ctcgcgcgga	tctttaaatg	6420
gagtgctctt	ttcccagttt	tcgcaatcca	catcggccag	atcgttattc	agtaagtaat	6480
ccaattcggc	taagcggctg	tctaagctat	tcgtataggg	acaatccgat	atgtcagatg	6540
agtgaagag	cctgtagcac	tcgcataca	gctcgataat	cttttcaggg	ctttgttctc	6600
cttcatactc	ttccgagcaa	aggacgccat	cggcctcact	catgagcaga	ttgtctcagc	6660
catcatgcgc	ttcaaagtgc	aggacctttg	gaacaggcag	ctttccttcc	agccatagca	6720
tcagtctctt	ttcccgttcc	acatcatagg	tggtcccttt	ataccggctg	tccgtcattt	6780
ttaaatatag	gttttcattt	tctcccacca	gcttatatac	cttagcagga	gacattcctt	6840
ccgtatcttt	tacgcagcgg	tatttttcga	tcagtttttt	caattccggt	gatattctca	6900
tttttagccat	ttattatttc	cttctctctt	tctacagtat	ttaaagatac	cccaagaagc	6960
taattataac	aagacgaact	ccaattcact	gttcccttga	ttctaaaacc	ttaaatacca	7020
gaaaacagct	ttttcaaagt	tgttttcaaa	gttggcggtat	aacatagtat	cgacggagcc	7080
gattttgaaa	ccgcgggtgat	cacaggcagc	aacgctctgt	catcgttaca	atcaacatgc	7140
tacctccgcg	gagatcatcc	gtgtttcaaa	cccgccagct	tagttgccgt	tcttccgaat	7200
agcatcggtg	acatgagcaa	agtctgcgcg	cttacaacgg	ctctcccgtc	gacgccgtcc	7260
cggactgatg	ggctgcctgt	atcgagtggg	gattttgtgc	cgagctgccg	gtcggggagc	7320
tggtggctgg	ctgggtggcag	gatatattgt	ggtgtaaaca	aattgacgct	tagacaactt	7380
aataacacat	tcgggacggt	tttaatgtac	tgaattaacg	ccgaattaat	tcgggggatc	7440
tggtttttag	tactggattt	tggttttagg	aattagaaat	tttattgata	gaagtatttt	7500
acaaatacaa	atacatacta	agggtttctt	atatgctcaa	cacatgagcg	aaaccctata	7560
ggaacctaa	ttcccttatc	tggaactac	tcacacatta	ttatggagaa	actcgagctt	7620
gtcgcagcac	agatccggct	ggcatctact	ctattctctt	gcccctcgac	gagtgctggg	7680
gcgtcgggtt	ccactatcgg	cgagtacttc	tacacagcca	tcgggtccaga	cggccgcgct	7740
tctgcggggc	atgtgtgtac	gcccgcagct	cccggtcccg	gatcggacga	ttgcctcgca	7800
tcgaccctgc	gcccagctgc	catcatcgaa	attgcccgtc	accaagctct	gatagagttg	7860
gtcaagaacca	atgcggagca	tatacgcccg	gagtcgtggc	gatcctgcaa	gctccgggat	7920
cctccgctcg	aagtgcgcg	tctgctgctc	catacaagcc	aaccacggcc	tccagaagaa	7980
gatgttggcg	acctcgatct	gggaatcccc	gaacatcgcc	tcgctccagt	caatgaccgc	8040
tggtatggcg	ccattgtccg	tcaggacatt	gttggagccg	aaatccgcgt	gcacgaggtg	8100
ccggacttcg	gggcagtcct	cggcccaaag	catcagctca	tcgagagcct	gcgcgacgga	8160
cgcactgaag	gtgtcgtcca	tcacagtttg	ccagtgtata	acatggggat	cagcaatcgc	8220
gcatatgaaa	tcacgccatg	tagtgtattg	accgatctct	tgccggtccga	atgggcccga	8280
cccgtctcgt	tggtcaagat	cggccgcagc	gatcgcatac	atagcctccg	cgaccggttg	8340
tagaacagcg	ggcagttcgg	tttcaggcag	gtcttgcaac	gtgacaccct	gtgcacggcg	8400
ggagatgcaa	taggtcaggc	tctcgctaaa	ctccccaatg	tcaagcactt	ccggaatcgg	8460
gagcggcgcc	gatgcaaaag	gccgataaac	ataacgatct	ttgtagaaac	catcggcgca	8520
gctattttacc	cgcaggacat	atccacgccc	tctcatatcg	aagctgaaag	cacgagattc	8580
ttcgcctctc	gagagctgca	tcaggctcga	gcagctgtcg	aacttttcga	tcagaaactt	8640
ctcgacagac	gtcgcgggtga	gttcaggctt	tttcatatct	cattgcccc	cgggatctgc	8700
gaaagctcga	gagagataga	ttttagtaga	gagactgggt	atttcagcgt	gtcctctcca	8760
aatgaaatga	acttccttat	atagagggaag	gtcttgcgaa	ggatagtggt	attgtgcgtc	8820
atcccttacg	tcagtgagga	tatcacatca	atccacttgc	tttgaagacg	tggttggaac	8880
gtcttctttt	tccacgatgc	tcctcgtggg	tgggggtcca	tctttgggac	cactgtcggc	8940
agaggcatct	tgaacgatag	cctttccttt	atcgcaatga	tggcatttgt	aggtgccacc	9000
ttcctttttc	actgtccttt	tgatgaagtg	acagatagct	gggcaatgga	atccgaggag	9060
gtttcccgat	attacccttt	gttgaagagt	ctcaatagcc	ctttggtctt	ctgagactgt	9120
atctttgata	ttctttggagt	agacgagagt	tcgctgctcc	accatgttat	cacatcaatc	9180
cacttgcttt	gaagacgtgg	ttggaacgtc	ttctttttcc	acgatgctcc	tcgtgggttg	9240
gggtccatct	ttgggaccac	tgctggcaga	ggcatcttga	acgatagcct	ttcctttatc	9300
gcaatgatgg	catttgtagg	tgccaccttc	cttttctact	gtccttttga	tgaagtgaca	9360

-53-

gatagctggg	caatggaatc	cgaggaggtt	tcccgatatt	accctttggt	gaaaagtctc	9420
aatagccctt	tggtcttctg	agactgtatc	tttgatattc	ttggagtaga	cgagagtgtc	9480
gtgctccacc	atgttggcaa	gctgctctag	ccaatacgca	aaccgcctct	ccccgcgctg	9540
tgcccgattc	attaatgcag	ctggcacgac	aggtttcccg	actggaaagc	gggcagttag	9600
cgcaacgcaa	ttaatgtgag	ttagctcact	cattaggcac	cccaggcttt	acactttatg	9660
cttccggctc	gtatgttgtg	tggaattgtg	agcggataac	aatttcacac	aggaaacagc	9720
tatgaccatg	attacaatt	cgagctcggg	accgggggat	cctctagagt	cgacctgcag	9780
gcctgcaagc	ttggcactgg	ccgtcgtttt	acaacgctgt	gactgggaaa	accctggcgt	9840
tacccaactt	aatcgccctt	cagcacatcc	ccctttcgcc	agctggcgta	atagcgaaga	9900
ggcccgccac	gatcgccctt	ccaacaggtt	gcccagcgctg	aatggcggaat	gctagagcag	9960
cttgagcttg	gatcagattg	tcgtttcccg	ccttcagttt	agcttcatgg	agtcaaagat	10020
tcaaataagag	gacctaacag	aactcgccgt	aaagactggc	gaacagttca	tacagagtct	10080
cttacgactc	aatgacaaga	agaaaatctt	cgtcaacatg	gtggagcacg	acacacttgt	10140
ctactccaaa	aatatcaaag	atacagtctc	agaagaccaa	agggcaattg	agacttttca	10200
acaaagggta	atatccggaa	acctcctcgg	attccattgc	ccagctatct	gtcactttat	10260
tgtgaagata	gtggaaaagg	aaggtggctc	ctacaaatgc	catcattgcg	ataaaggaaa	10320
ggccatcggt	gaagatgcct	ctgccgacag	tggtocccaa	gatggacccc	caccacgag	10380
gagcatcggt	gaaaaagaag	acgttccaac	cacgtcttca	aagcaagtgg	attgatgtga	10440
tatctccact	gacgtaagg	atgacgcaca	atcccactat	ccttcgcaag	acccttcctc	10500
tatataagga	agttcatttc	atttggagag	aacacggggg	actcttgac		10549

<210> 93

<211> 33

<212> DNA

<213> Artificial Sequence

<220>

<223> CaMV35SpolyA Primer

<400> 93

ctgaattaac gccgaattaa ttcgggggat ctg 33

<210> 94

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> CaMV35Spr Primer

<400> 94

ctagagcagc ttgccaacat ggtggagca 29

<210> 95

<211> 12592

<212> DNA

<213> Artificial Sequence

<220>

<223> pAg2 Plasmid

<400> 95

gtacgaagaa	ggccaagaac	ggccgcctgg	tgacgggtatc	cgagggtgaa	gccttgabta	60
gcccgtacaa	gatcgtaaa	agcgaaaccg	ggcgcccgga	gtacatcgag	atcgagctag	120
ctgattggat	gtaccgcgag	atcacagaag	gcaagaaccc	ggacgtgctg	acggttcacc	180
ccgattactt	tttgatcgat	cccggcatcg	gcccgtttct	ctaccgcctg	gcacgcccg	240
ccgcaggcaa	ggcagaagcc	agatgggtgt	tcaagacgat	ctacgaacgc	agtggcagcg	300
ccggagagtt	caagaagttc	tgtttcaccc	tgcgcaagct	gatcgggtca	aatgacctgc	360
cggagtacga	tttgaaggag	gaggcggggc	aggctggccc	gatcctagtc	atgcgtacc	420
gcaacctgat	cgaggggcaa	gcatccggcc	gttcctaata	tacggagcag	atgctagggc	480
aaattgccct	agcaggggaa	aaaggtcgaa	aaggctctct	tcctgtggat	agcacgtaca	540
ttgggaaccc	aaagccgtac	attgggaacc	ggaacccgta	cattgggaac	ccaaagccgt	600
acattgggaa	ccggtcacac	atgtaagtga	ctgatataaa	agagaaaaaa	ggcgattttt	660
ccgcctaaaa	ctctttaaaa	cttattaaaa	ctcttaaaac	ccgcctggcc	tgtgcataac	720
tgtctggcca	gcgcacagcc	gaagagctgc	aaaaagcgcc	tacccttcgg	tcgctgcgct	780
ccctacggcc	cgccgcttcg	cgtcggccta	tcgcgccgcg	tgcccgctca	aaaatggctg	840
gcctacggcc	aggcaatcta	ccagggcgcg	gacaagccgc	gccgtcgcca	ctcgaccgcc	900

-54-

ggcgcccaaca	tcaaggcacc	ctgcctcgcg	cgtttcggtg	atgacgggtga	aaacctctga	960
cacatgcagc	tcccggagac	ggtcacagct	tgtctgtaag	cggatgcccgg	gagcagacaa	1020
gcccgtcagg	gcgcgtcagc	gggtgttgcc	gggtgtcggg	gocgagccat	gaccagtcga	1080
cgtagcgata	gcggagtgtg	tactggctta	actatgcggc	atcagagcag	attgtactga	1140
gagtgcacca	tatgcgggtg	gaaataccgc	acagatgcgt	aaggagaaaa	taccgcatca	1200
ggcgctcttc	cgcttccctg	ctcactgact	cgctgcgctc	ggtcggttcgg	ctgcggcgag	1260
cgggtacagc	tcactcaaag	gcggtaatac	ggttatccac	agaatcaggg	gataacgcag	1320
gaaagaacat	gtgagcaaaa	ggccagcaaa	agggcaggaa	ccgtaaaaag	gccgcgttgc	1380
tggcggtttt	ccataggctc	cgccccctg	acgagcatca	caaaaaatcga	cgctcaagtc	1440
agagggtggcg	aaacccgaca	ggactataaa	gataccaggc	gtttccccct	ggaagctccc	1500
tcgtgcgctc	tcctgttccg	accctgccgc	ttaccggata	cctgtccgcc	tttctccctt	1560
cgggaagcgt	ggcgctttct	catagctcac	cctgtaggta	tctcagttcg	gtgtaggctcg	1620
ttcgctccaa	gctgggctgt	gtgcacgaac	gccccgttca	gcccagaccgc	tgcgccttat	1680
cggtaacta	tcgtcttgag	tccaacccgg	taagacacga	cttatcgcca	ctggcagcag	1740
ccactggtaa	caggattagc	agagcgaggt	atgtaggcgg	tgctacagag	ttcttgaagt	1800
ggtggcctaa	ctacggctac	actagaagga	cagtatttgg	tatctgcgct	ctgctgaagc	1860
cagttacott	cggttaacgt	gttggtagct	cttgatccgg	caaacaaaacc	accgctggta	1920
gcgggtggtt	ttttgtttgc	aagcagcaga	ttacgcgcag	aaaaaaagga	tctcaagaag	1980
atcctttgat	cttttctacg	gggtctgacg	ctcagtgga	cgaaaaactca	cgtaaagggg	2040
ttttggtcac	gcattctagg	tactaaaaca	attcatccag	taaaatataa	tattttattt	2100
tctcccaatt	aggcttgatc	cccagtaagt	caaaaaatag	ctcgacatac	tgttctgtcc	2160
cgatatcctc	cctgatcgac	cggacgcaga	aggcaatgtc	ataccacttg	tccgccttgc	2220
cgcttctccc	aagatcaata	aagccactta	ctttgccatc	tttcacaaag	atgttgctgt	2280
ctcccaggtc	gccgtgggaa	aagacaagtt	cctcttcggg	cttttccgct	tttaaaaaat	2340
catcacgctc	gcgcggatct	ttaaatggag	tgtcttcttc	ccagttttcg	caatccacat	2400
cggccagatc	gttattcagt	aagtaatcca	attcggctaa	gcggctgtct	aagctattcg	2460
tatagggaca	atccgatatg	tcgatggagt	gaaagagcct	gatgcactcc	gcatacagct	2520
cgataatctt	ttcagggtct	tgttcatctt	catactcttc	cgagcaaagg	acgccatcgg	2580
cctcactcat	gagcagattg	ctccagccat	catgccgttc	aaagtgcagg	acctttggaa	2640
caggcagctt	tccttccagc	catagcatca	tgtccttttc	ccgttccaca	tcatagggtg	2700
tccctttata	cggctgttcc	gtcattttta	attcggctaa	ttcattttct	cccaccagct	2760
tatatacctt	agcaggagac	attccttccg	tatcttttac	gcagcgggat	ttttcgatca	2820
gttttttcaa	ttccgggtgat	attctcattt	tagccattta	ttatttcctt	cctcttttct	2880
acagtattta	aagatacccc	aagaagctaa	ttataacaag	acgaactcca	attcactgtt	2940
ccttgacttc	taaaacctta	aataccagaa	aacagctttt	tcaaagttgt	tttcaaagtt	3000
ggcgataaac	atagatcga	cggagccgat	tttgaaaacc	cgggtgatcac	aggcagcaac	3060
gctctgtcat	cgttacaatc	aacatgctac	cctccgcgag	atcatccgtg	tttcaaacc	3120
ggcagcttag	ttgccgttct	tccgaatagc	atcggtaaca	tgagcaaagt	ctgccgcctt	3180
acaacggctc	tcccgtgtac	gccgtcccgg	actgatgggc	tgccctgtatc	gagtgggtgat	3240
tttgtccga	tgctccggtc	ggggagctgt	tggtctggctg	gtggcaggat	atattgtgtg	3300
gtaaacaaat	tgacgccttag	acaacttaat	aacacattgc	ggacgttttt	aatgtactga	3360
attaacgccg	aattaattcg	ggggatctgg	atttttagtac	tggatttttg	ttttaggaat	3420
tagaaatttt	attgatagaa	gtattttaca	aatacaataa	catactaagg	gtttcttata	3480
tgctcaaac	atgagcga	ccctatagga	accctaattc	ccttatcttg	gaactactca	3540
cacatttata	tggagaact	cgagtcaaat	cctcgtgacg	ggcaggaccg	gacggggcgg	3600
taccggcagg	ctgaagtcca	gctgccagaa	acccacgtca	tgccagttcc	cgtgcttgaa	3660
gocggccgcc	cgcagcatgc	cgcggggggc	atatccgagc	gcctcgtgca	tgcgcacgct	3720
cgggtcgttg	ggcagcccga	tgacagcgac	cacgctcttg	aagccctgtg	cctccaggga	3780
cttcagcagg	tggtgttaga	gcgtggagcc	cgctcccgct	cgctgggtggc	ggggggagac	3840
gtacacggtc	tccagtccga	tccagtccga	ggcgttgctg	gccttccagg	ggcccgcgta	3900
ggcgatgccc	gcgacctcgc	cgtccacctc	ggcgacgagc	cagggatagc	gctcccgcag	3960
acggacgagg	tcgtccgctc	actcctgcgg	ttcctgcggc	tcggtagcga	agttgaccgt	4020
gcttgtctcg	atgtagtgg	tgacgatgg	gcagacccgc	ggcatgtccg	cctcgggtggc	4080
acggcggatg	tcggccgggc	gtcgttctgg	gctcatggta	gactcgagag	agatagattt	4140
gtagagagag	actgggtgatt	tcagcgtgtc	ctctccaaat	gaaatgaact	tccttatata	4200
gaggaagggtc	ttgcgaagga	tagtgggatt	gtgcgtcatc	ccttacgtca	gtggagatat	4260
cacatcaatc	cacttgcttt	gaagacgtgg	ttggaacgtc	ttctttttcc	acgatgctcc	4320
tcgtgggttg	gggtccatct	ttgggaccac	tgtcggcgaga	ggcatcttga	acgatagcct	4380
ttcctttatc	gcaatgatgg	catttgtagg	tgccaccttc	cttttctact	gtccttttga	4440
tgaagtgaca	gatagctggg	caatgggaatc	cgaggagggt	tcccgatatt	accttttgtt	4500
gaaaagtctc	aatagccctt	tggtcttctg	agactgtatc	tttgatattc	ttggagtaga	4560
cgagagtgtc	gtgctccacc	atgttatcac	atcaatccac	ttgctttgaa	gacgtgggtg	4620
gaacgtcttc	tttttccacg	atgctcctcg	tgggtggggg	tccatctttg	ggaccactgt	4680
cggcagagge	atcttgaaacg	atagcctttc	ctttatcgca	atgatggcat	ttgtaggctg	4740
caccttccct	ttctactgtc	cttttgatga	agtgacagat	agctgggcaa	tggaaatccga	4800
ggaggtttcc	cgatattacc	ctttgttgaa	aagtctcaat	agccctttgg	tcttctgaga	4860
ctgtatcttt	gatattcttg	gagtagacga	gagtgtcgtg	ctccaccatg	ttggcaagct	4920

-55-

gctctagcca	atacgcaaac	cgcctctccc	cgcgcgttgg	ccgattcatt	aatgcagctg	4980
gcacgacagg	tttcccgaact	ggaaagcggg	cagtgcgcgc	aacgcaatta	atgtgagtta	5040
gctcactcat	taggcacccc	aggctttaca	ctttatgctt	ccggctcgta	tggtgtgtgg	5100
aattgtgagc	ggataacaat	ttcacacagg	aaacagctat	gaccatgatt	acgaattcga	5160
gccttgacta	gaggggtcgac	ggtatacaga	catgataaga	tacattgatg	agttttggaca	5220
aaccacaact	agaatgcagt	gaaaaaaatg	ctttatttgt	gaaatttgtg	atgctattgc	5280
tttatttcta	accattataa	gctgcaataa	acaagttggg	gtgggcgaag	aactccagca	5340
tgagatcccc	gcgctggagg	atcatccagc	cggcgtcccc	gaaaaacgatt	ccgaagccca	5400
acctttcata	gaaggcggcg	gtggaatcga	aatctcgtag	cacgtgtcag	tcctgtctct	5460
cggccacgaa	gtgcacgcag	ttgccggccg	ggtcgcgcag	ggcgaactcc	cgccccacg	5520
gctgctcgcc	gatctcggtc	atggccggcc	cggaggcgtc	ccggaagtcc	gtggacacga	5580
cctccgacca	ctcggcgtag	agctcgtcca	ggcgcgcac	ccacacccag	gccagggtgt	5640
tgctcggcac	cactcgggtc	tggaccgcgc	tgatgaacag	ggtcacgtcg	tcccggacca	5700
caccggcgaa	gtcgtcctcc	acgaagtccc	gggagaaccc	gagccggtcg	gtccagaact	5760
cgaccgtctc	ggcgagctcg	cgcgcgggtga	gcaccggaac	ggcactgggtc	aacttggcca	5820
tggatccaga	tttcgctcaa	gtagtataa	aaaagcaggc	ttcaatcctg	caggaattcg	5880
atcgacactc	tcgtctactc	caagaatata	aaagatacag	tctcagaaga	ccaaaggggt	5940
attgagactt	ttcaacaaag	ggtaatatcg	ggaaacctcc	tcggattcca	ttgccagct	6000
atctgtcact	tcatcaaaag	gacagtagaa	aagggaaggtg	gcacctacaa	atgccatcat	6060
tgcgataaag	gaaaggctat	cgttcaagat	gcctctgccc	acagtgggtcc	caaagatgga	6120
ccccaccaca	cgaggagcat	cgtggaaaaa	gaagacgttc	caaccacgtc	ttcaaagcaa	6180
gtggattgat	gtgataacat	ggtggagcac	gacactctcg	tctactccaa	gaatatcaaa	6240
gatacagtct	cagaagacca	aagggtctatt	gagacttttc	aacaaagggg	aatatcgggg	6300
aacctcctcg	gattccattg	cccagctatc	tgctcacttca	tcaaaaggac	agtagaaaag	6360
gaaggtggca	cctacaaaatg	ccatcattgc	gataaaaggaa	aggctatcgt	tcaagatgcc	6420
tctgcccaga	gtgggtcccaa	agatggaccc	ccaccacga	ggagcatcgt	ggaaaaagaa	6480
gacgttccaa	ccagctcttc	aaagcaagtg	gattgatgtg	atatctccac	tgacgtaagg	6540
gatgacgcac	aatccacta	tccttcgcaa	gaccttctct	tatataagga	agttcatttc	6600
atgtggagag	gacacgctga	aatcaccagt	ctctctctac	aaatctatct	ctctcgagct	6660
ttcgagatc	cgggggggca	atgagatatg	aaaaagcctg	aactcaccgc	gacgtctgtc	6720
gagaagtctc	tgatcgaaaa	gttcgacagc	gtctccgacc	tgatgcagct	ctcggagggc	6780
gaagaatctc	gtgcttttcag	tttcgatgta	ggagggcgtg	gatattgtct	gcgggtaaat	6840
agctgcgccc	atgggtttcta	caaagatcgt	tatgtttatc	ggcacttttg	atcggcccg	6900
ctcccgaattc	cggaagtgtc	tgacattggg	gagtttagcg	agagcctgac	ctattgcatc	6960
tcccgcgctg	cacagggtgt	cacgttgcaa	gacctgcctg	aaaccgaact	gcccgcgtgt	7020
ctacaaccgg	tcgcccaggc	tatggatgcg	atcgtcgcgg	ccgatcttag	ccagacgagc	7080
gggttcggcc	cattccgacc	gcaaggaatc	ggtcaataca	ctacatggcg	tgatttcata	7140
tgccgcgattg	ctgatcccca	tgtgtatcac	tggcaaaactg	tgatggacga	caccgtcagt	7200
gcgtccgtcg	cgcaggctct	cgatgagctg	atgctttggg	ccgaggactg	ccccgaagtc	7260
cggcacctcg	tgcacgcgga	tttcggctcc	aaacatgtcc	tgacggacaa	tggccgcata	7320
acagcggtca	ttgactggag	cgaggcgatg	ttcggggatt	cccaatacga	gggtcgccaac	7380
atcttcttct	ggaggccgtg	gttggcttgt	atggagcagc	agacgcgcta	cttcgagcgg	7440
aggcatccgg	agcttgcagg	atcgccacga	ctccgggcgt	atatgctccg	cattgggtct	7500
gaccaactct	atcagagctt	ggttgacggc	aatttcgatg	atgcagcttg	ggcgcaggct	7560
cgatgcgcag	caatcgtccg	atccggagcc	ggagactgtc	ggcgtacaca	aatcgcccg	7620
agaagcgcgg	ccgtctggac	cgatgcgtgt	tgagaagtac	tcgccgatag	tggaaaccga	7680
cgccccagca	ctcgtccgag	ggcaaagaaa	tagagtagat	gccgaccgga	tctgtcgatc	7740
gacaagctcg	agttttctcca	taataatgtg	tgagtgttct	ccagataagg	gaattagggg	7800
tctctatagg	tttcgctcat	gtgttgagca	tataagaaac	ccttagtatg	tatttgtatt	7860
tgtaaaatac	ttctatcaat	aaaatttcta	attcctaata	ccaaaatcca	gtactaaaat	7920
ccagatcccc	cgaattaatt	cggcgttaat	tcagatcaag	cttggcactg	gcctgcgttt	7980
tacaacgtcg	tgactgggaa	aaccctggcg	ttaccacaact	taatcgctt	gcagcacatc	8040
cccccttcgc	cagctggcgt	aatagcgaag	aggcccgac	cgatcgccct	tcccaacagt	8100
tgcgcagcct	gaattggcgaa	tgttagagca	gcttgagctt	ggatcagatt	gtcgtttccc	8160
gccttcaggt	tgggggattcct	ctagactgaa	ggcgggaaac	gacaatctga	tcatgagcgg	8220
agaattaagg	gagtcacgtt	atgacccccg	ccgatgacgc	gggacaagcc	gttttacgtt	8280
tggaactgac	agaaccgcaa	cgttgaagga	gccactcagc	cgcgggtttc	tggagtttaa	8340
tgagctaagc	acatacgtca	gaaaccatta	ttgcgcgttc	aaaagtgcgc	taaggctcat	8400
atcagctagc	aaataattct	tgtcaaaaat	gctccactga	cgttccataa	attccctcgc	8460
gtatccaatt	agagtctcat	attcaactct	aatccaaata	atctgcaccg	gatctcgaga	8520
atcgaattcc	cgcggccgcg	atggtagatc	tgactagtaa	aggagaagaa	cttttctactg	8580
gagttgtccc	aattcttgtt	gaattagatg	gtgatgttaa	tgggcacaaa	ttttctgtca	8640
gtggagaggg	tgaagggtgat	gcaacatac	gaaaacttac	ccttaaattt	atttgcacta	8700
ctggaaaaact	acctgttccg	tggccaacac	ttgtcactac	ttctcttctt	gggtgttcaa	8760
gcttttcaag	ataccagat	catatgaagc	ggcacgactt	cttcaagagc	gccaatgcctg	8820
agggatacgt	gcaggagagg	accatcttct	tcaaggacga	cgggaactac	aagacacgtg	8880
ctgaagtcaa	gtttgagggg	gacacccctc	tcaacaggat	cgaacttaag	ggaatcgatt	8940

-56-

tcaaggagga	cggaaacatc	ctcgccacac	agttggaata	caactacaac	tcccacaacg	9000
tatacatcat	ggccgacaag	caaaaagaacg	gcatcaaagc	caacttcaag	acccgcccac	9060
acatcggaaga	cggcggcgtg	caactcgctg	atcattatca	acaaaatact	ccaattggcg	9120
atggccctgt	cctttttacca	gacaaccatt	acctgtccac	acaatctgcc	ctttcgaaag	9180
atcccaacga	aaagagagac	cacatgggtcc	ttcttgagtt	tgtaacagct	gctgggatta	9240
cacatggcat	ggatgaacta	tacaaagcta	gccaccacca	ccaccaccac	gtgtgaattg	9300
gtgaccagct	cgaatttccc	cgatcgttca	aacatttggc	aataaagttt	cttaagattg	9360
aatcctgttg	cgggtcttgc	gatgattatc	atataatttc	tgttgaatta	cgtaaagcat	9420
gtaataatta	acatgtaatg	catgacgtta	tttatgagat	gggtttttat	gattagagtc	9480
ccgcaattat	acattttaata	cgcgatagaa	aacaaaatat	agcgcgcaaa	ctaggataaaa	9540
ttatcgcgcg	cgggtgtcatc	tatgttacta	gatcgggaaat	taaactatca	gtgttttgaca	9600
ggatataattg	gcgggttaaac	ctaagagaaa	agagcgttta	ttagaataac	ggatatttaa	9660
aaagggcgtga	aaaggttttat	ccgttcgtcc	atltgtatgt	gcatgccaac	cacaggggttc	9720
ccctcgggat	caaagtacttt	tgatccaacc	cctccgctgc	tatagtgcag	tcggcttctg	9780
acgttcagtg	cagccgtctt	ctgaaaacga	catgtcgcac	aagtcctaag	ttacgcgcaca	9840
ggctgcgcgc	ctgccctttt	cctggcgttt	tcttgtcgcg	tgttttagtc	gcataaagta	9900
gaataacttg	gactagaacc	ggagacatta	gccatgaac	aagagcgccg	ccgctggcct	9960
gctgggctat	gcccgcgtca	gcaccgacga	ccaggacttg	accaaaccaac	gggcccgaact	10020
gcacgcggcc	ggctgcacca	agctgttttc	cgagaagatc	accggcacca	ggcgcgaccg	10080
cccggagctg	gccaggatgc	ttgaccacct	acgccctggc	gacgttgtga	cagtgcaccag	10140
gctagaccgc	ctggcccgcga	gcaccgcgca	cctactggac	attgccgagc	gcatccagga	10200
ggcgcgcgcg	ggcctgcgta	gcctggcaga	gcctgtggcc	gacaccacca	cgccggccgg	10260
ccgcatgggtg	ttgaccgtgt	tcgcccgcac	tgccgagttc	gagcgttccc	taatcatcga	10320
ccgcaccccg	agcgggcgcg	aggccgccaa	ggcccagagg	gtgaagtttg	gcccccgccc	10380
taccctcacc	ccggcacaga	tcgcccgcgc	ccgcgagctg	atcgaccagg	aaggccgcac	10440
cgtgaaagag	gcggctgcac	tgcttggcgt	gcatcgctcg	accctgtacc	gcgcacttga	10500
gcgcagcgag	gaagtgcagc	ccaccgaggc	caggcggcgc	ggtgccttcc	gtgaggacgc	10560
attgaccgag	gccgacgccc	tgccggccgc	cgagaatgaa	cgccaagagg	aacaagcatg	10620
aaaccgcacc	aggacggcca	ggacgaaccg	tttttcatta	ccgaagagat	cgaggcggag	10680
atgatcgcg	ccgggtacgt	gttcgagccg	cccgcgcacg	tctcaaccgt	gcggctgcac	10740
gaaatcctgg	ccggtttgtc	tgatgccaaag	ctggcggcct	ggccggccag	cttggccgct	10800
gaagaaaccc	agcgcgcgcg	tctaaaaagg	tgatgtgtat	ttgagtaaaa	cagcttgcgt	10860
catgcggtcg	ctgcgtatat	gatgcgatga	gtaaataaac	aaatacgcaa	ggggaaacgca	10920
tgaaggttat	cgctgtactt	aaccagaaaag	gcgggtcagg	caagacgacc	atcgcaaccc	10980
atctagcccg	cgccctgcaa	ctgcgcgggg	ccgatgttct	gttagtcgat	tcogatacccc	11040
agggcagtcg	cgcgcgattg	gcggccgtgc	gggaagatca	accgctaacc	gttgtcggca	11100
tcgaccgccc	gacgattgac	cgcgacgtga	aggccatcgg	ccggcgcgcac	ttcgtagtga	11160
tcgacggagc	gcccacggcg	gcggacttgg	ctgtgtccgc	gatcaaggca	gccgacttcg	11220
tgctgattcc	ggtgcagcca	agcccttacg	acatatgggc	caccgcgcgac	ctggtggagc	11280
tggttaagca	gcgcattgag	gtcagggatg	gaaggctaca	agcggccttt	gtcgtgtcgc	11340
gggcgatcaa	aggcacgcgc	atcggcggtg	aggttgccga	ggcgtgggcc	gggtacgagc	11400
tgcccattct	tgagtccgt	atcacgcagc	gcgtgagcta	cccaggcact	gccgcgcgcg	11460
gcacaaccgt	tcttgaatca	gaaccgcagg	gcgacgctgc	ccgcgaggtc	caggcgctgg	11520
ccgctgaaat	taaatacaaaa	ctcattttgag	ttaatgaggt	aaagagaaaa	tgagcaaaaag	11580
cacaacacag	ctaagtgcgc	gccgtccgag	cgcacgcagc	agcaaggctg	caacgttggc	11640
cagcctggca	gacacgccag	ccatgaagcg	ggtcaacttt	cagttgccgg	cggaggatca	11700
caccaagctg	aagatgtacg	cggtagccca	aggcaagacc	attaccgagc	tgctatctga	11760
atacatcgcg	cagctaccag	agtaaatgag	caaatgaata	aatgagtaga	tgaatttttag	11820
cggctaaagg	aggcggcatg	gaaaaatcaag	aacaaccagg	caccgacgcc	gtggaatgcc	11880
ccatgtgtgg	aggaacgggc	ggttggccag	gcgtaagcgg	ctgggttgtc	tgccggccct	11940
gcaatggcac	tggaaacccc	aagcccaggg	aatcggcgtg	acggtcgcaa	accatccggc	12000
ccggtacaaa	tcggcgcggc	gctgggtgat	gacctggtgg	agaagttgaa	ggccgcgcgag	12060
gccgcccagc	ggcaacgcac	cgaggcagaa	gcacgcccgc	gtgaatcgctg	gcaagcggcc	12120
gctgatcgaa	tccgcacaa	atcccggcaa	ccgcggcgag	ccggtgcgcc	gtcgattagg	12180
aaagccgcca	agggcgacga	gcaaccagat	tttttcgttc	cgatgctcta	tgacgtgggc	12240
accgcgcgata	gtcgcagcat	catggacgtg	gccgtttttc	gtctgtcgaa	gcgtgaccga	12300
cgagctggcg	aggtgatccg	ctacgagctt	ccagacgggc	acgtagaggt	ttccgcaggg	12360
ccggccggca	tggccagtg	gtgggattac	gacctggtac	tgatggcgg	ttcccatcta	12420
accgaatcca	tgaaccgata	ccgggaagg	aagggagaca	agcccggccg	cgtgtccgt	12480
ccacacgttg	cggacgtact	caagttctgc	cggcgagccg	atggcggaaa	gcagaaagac	12540
gacctggtag	aaacctgcat	tcggttaaac	accacgcacg	ttgccatgca	gc	12592

<210> 96

<211> 3357

<212> DNA

<213> Artificial Sequence

-57-

<220>

<223> pGEMEasyNOS Plasmid

<400> 96

tatcactagt	gaattcgcgg	ccgcctgcag	gtcgaccata	tgggagagct	cccaacgcgt	60
tggatgcata	gcttgagtat	tctatagtg	cacctaata	gcttggcgta	atcatgggtca	120
tagctgtttc	ctgtgtgaaa	ttgttatccg	ctcaccaattc	cacacaacat	acgagccgga	180
agcataaagt	gtaaagcctg	gggtgcctaa	tgaagtgcagct	aactcacatt	aattgcgttg	240
cgctcactgc	ccgcttttcca	gtcgggaaac	ctgtcgtgcc	agctgcatta	atgaatcggc	300
caacgcgcgg	ggagaggcgg	tttgcgattt	ggcgctctct	ccgcttcctc	gctcactgac	360
tcgctgcgct	cggtcggttcg	gctgcggcga	gcggtatcag	ctcactcaaa	ggcggtaata	420
cggttatcca	cagaatcagg	ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	480
aaggccagga	accgtaaaaa	ggccgcgttg	ctggcgtttt	tccatagggt	ccgccccct	540
gacgagcatc	acaaaaatcg	acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	600
agataccagg	cgtttccccc	tggaaagctcc	ctcgtgcgct	ctcctgttcc	gaccctgccg	660
cttacccgat	acctgtccgc	ctttctccct	tcgggaagcg	tggcgcttcc	tcatagctca	720
cgctgtaggt	atctcagttc	ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	780
ccccccgttc	agcccgaccg	ctgcgcctta	tccggtaact	atcgtcttga	gtccaacccg	840
gtaagacacg	acttatcgcc	actggcagca	gccaactggt	acaggattag	cagagcgagg	900
tatgtaggcg	gtgctacaga	gttcttgaag	tgggtggccta	actacggcta	cactagaaga	960
acagtatattg	gtatctgcgc	tctgctgaag	ccagttacct	tcggaaaaag	agttggtagc	1020
tcttgatccg	gcaaacaaaac	caccgctggt	agcgggtggt	tttttgtttg	caagcagcag	1080
attacgcgca	gaaaaaaaagg	atctcaagaa	gatcctttga	tcttttctac	ggggctctgac	1140
gctcagtgga	acgaaaactc	acgttaaggg	attttgggtca	tgagattatc	aaaaaggatc	1200
ttcacctaga	tcctttttaa	ttaaaaatga	agttttaaat	caatctaaag	tatatatgag	1260
taaacttggt	ctgcagttta	ccaatgcctta	atcagtgagg	cacctatctc	agcagatctgt	1320
ctatttcggt	catccatagt	tgctgactc	cccgctggtg	agataactac	gatacgggag	1380
ggcttaccat	ctggccccag	tgctgcaatg	ataccgcgag	acccacgctc	accggctcca	1440
gatttatcag	caataaacca	gccagccgga	agggccgagc	gcagaagtgg	tcctgcaact	1500
ttatccgcct	ccatccagtc	tattaattgt	tgccgggaag	ctagagtaag	tagttcgcca	1560
gttaatatgtt	tgcgcaacgt	tgttgccatt	gctacaggca	tcgtggtgtc	acgctcgctc	1620
tttggtatgg	cttcatttcag	ctccggttcc	caacgatcaa	ggcgagttac	atgatcccc	1680
atgttgtgca	aaaaagcggg	tagctccttc	ggtcctccga	tcgttgtcag	aagtaagttg	1740
gccgcagtg	tatcactcat	ggttatggca	gcactgcata	attctcttac	tgtcatgcc	1800
tcogtaagat	gcttttctgt	gactggtgag	tactcaacca	agtcattctg	agaatatgtg	1860
atgcggcgac	cgagttgctc	ttgcccggcg	tcaatacggg	ataataccgc	gccacatagc	1920
agaactttaa	aagtgtcat	cattggaaaa	cgttcttcgg	ggcgaaaact	ctcaaggatc	1980
ttaccgctgt	tgagatccag	ttcgatgtaa	cccactcggt	cacccaactg	atcttcagca	2040
tccttttactt	tcaccagcgt	ttctgggtga	gcaaaaaacag	gaaggcaaaa	tgccgcaaaa	2100
aagggaataa	gggcgacacg	gaaatgttga	atactcatat	tcttcccttt	tcaatatatt	2160
tgaagcattt	atcagggtta	tgtctcatg	agcggataca	tatttgaatg	tatttagaaa	2220
aataaacaaa	taggggttcc	gcgcacattt	ccccgaaaag	tgccacctga	tcgggtgtga	2280
aataccgcac	agatgcgtaa	ggagaaaaata	ccgcatcagg	aaattgtaag	cgttaatatt	2340
ttgttaaaaa	tcgcgttaaa	tttttgttaa	atcagctcat	tttttaacca	ataggccgaa	2400
atccgcaaaa	tcctttataa	atcaaaaagaa	tagaccgaga	taggggttgag	tgttgttcca	2460
gtttggaaca	agagtccact	attaaagaac	gtggactcca	acgtcaaagg	gcgaaaaacc	2520
gtctatcagg	gcgatggccc	actacgtgaa	ccatcaccct	aatcaagttt	tttggggctg	2580
agggtgccgt	aagcactaaa	tcggaaccct	aaagggagcc	cccgatttag	agcttgacgg	2640
ggaaagccgg	cgaacgtggc	gagaaaggaa	gggaagaaag	cgaagggagc	gggcgctagg	2700
gcgctggcaa	gtgtagcggg	cacgctgcgc	gtaaccacca	cacccgcgcg	gcttaattgc	2760
ccgctacagg	gcgcgtccat	tcgccattca	ggctgcgcaa	ctgttgggaa	gggcgcatcg	2820
tgccgggcctc	ttcgctatta	cgccagctgg	cgaagggggg	atgtgctgca	aggcgattaa	2880
ggtgggtaac	gccagggttt	tcccagtcac	gacgttgtaa	aacgacggcc	agtgaattgt	2940
aatacgactc	actatagggc	gaattgggccc	cgacgtcgca	tgctcccggc	cgccatggcg	3000
gccgcgggaa	ttcgattctc	gagatccggg	cgagattatt	tggattgaga	gtgaatatga	3060
gactctaatt	ggataccgag	gggaatttat	ggaacgtcag	tggagcattt	ttgacaagaa	3120
atatttgcta	gctgatagtg	accttaggcg	acttttgaac	gcgcaataat	ggtttctgac	3180
gtatgtgctt	agctcattaa	actccagaaa	cccgcggctg	agtggctcct	tcaacgttgc	3240
ggttctgtca	gttccaaacg	taaaacgggt	tgtcccgcgt	catcggcggg	ggtcataacg	3300
tgactccctt	aattctccgc	tcattgatcag	attgtcgttt	cccgccttca	gtctaga	3357

<210> 97

<211> 10122

<212> DNA

<213> Artificial Sequence

<220>

-58-

<223> p1302NOS Plasmid

<400> 97

```

catggttagat ctgactagta aaggagaaga acttttctact ggagttgtcc caattcttgt 60
tgaattagat ggtgatgtta atgggcacaa attttctgtc agtggagagg gtgaagggtga 120
tgcaacatac ggaaaactta cccttaaatt tatttgcact actggaaaac tacctgttcc 180
gtggccaaca ctgtcacta ctttctctta tgggtgtcaa tgcttttcaa gataccaga 240
tcatatgaag cggcacgact tcttcaagag cgccatgcct gagggatacg tgcaggagag 300
gaccatcttc ttcaaggacg acgggaacta caagacacgt gctgaagtca agtttgaggg 360
agacaccctc gtcaacagga tcgagcttaa gggaatcgat ttcaaggagg acggaaacat 420
cctcgccac aagttggaat acaactacaa ctcccacaac gtatacatca tggccgacaa 480
gcaaaagAAC ggcatcaaag ccaacttcaa gaccgcac aacatcgaag acggcgcg 540
gcaactcgct gatcattatc aacaaaatac tccaattggc gatggccctg tccttttacc 600
agacaaccat tacctgtcca cacaatctgc ctttgcgaaa gatcccaacg aaaagagaga 660
ccacatggtc cttcttgagt ttgtaacagc tgctgggatt acacatggca tggatgaact 720
atacaaagct agccaccacc accaccacca cgtgtgaatt ggtgaccagc tcgaatttcc 780
ccgatcgctt aaacatttgg caataaagtt tcttaagatt gaatcctggt gccggtcttg 840
cgatgattat catataattt ctggtgaatt acgttaagca tgtaataatt aacatgtaat 900
gcatgacgtt atttatgaga tgggttttta tgattagagt cccgcaatta tacattttaat 960
acgcgataga aaacaaaata tagcgcgcaa actaggataa attatcgcg gccggtgtcat 1020
ctatgttact agatcgggaa ttaactatc agtgtttgac aggatataat ggccgggtaaa 1080
cctaagagaa aagagcgttt attagaataa cggatattta aaagggcggtg aaaagggtta 1140
tccgttcgtc catttgatg tgcatgccaa ccacagggtt cccctcgga tcaaagtact 1200
ttgatccaac ccctccgctg ctatagtgc gtccggttct gacgttcagt gcagccgtct 1260
tctgaaaacg acatgtcgca caagtcctaa gttacgcgac aggtgcccgc cctgcccctt 1320
tcctggcggt ttctgtcgcg gtgttttagt cgcataaagt agaatacttg cgactagaac 1380
cggagacatt acgccatgaa caagagcgcc gccctggcc tgctgggcta tgcccgctc 1440
agcaccgacg accaggactt gaccaaccaa cgggccgaac tgcacgcggc cggctgcacc 1500
aagctgtttt ccgagaagat caccggcacc aggcgcgacc gcccgagct ggccaggatg 1560
cttgaccacc tacgcccctg cgacgttgtg acagtgaaca ggctagaccg cctggcccgc 1620
agcacccgcg acctactgga cattgccgag cgcattccagg aggcggcgcc gggcctgcgt 1680
agcctggcag agcctggggc cgacaccacc agcccgccg gccgcatggt gttgaccgtg 1740
ttcgccggca ttgcccaggt cgagcggttc ctaatcatcg accgcacccg gagcgggcg 1800
gaggccgcca aggcggcgag cgtgaagttt ggcccccgcc ctaccctcac cccggcacag 1860
atcgcgacg cccgcgagct gatcgaccag gaagccgca ccgtgaaaga ggccggtgca 1920
ctgcttggcg tgcctcgctc gacctgtac cgcgacttg agcgacgca ggaagtgcg 1980
cccaccgagg ccaggcgggc cgggtgcctt cgtgaggacg cattgaccga ggccgacgcc 2040
ctggcgccg ccgagaatga acgccaagag gaacaagcat gaaaccgcac caggacggcc 2100
aggacgaacc gtttttcatt accgaagaga tcgaggcgga gatgatcgcg gccgggtacg 2160
tgttcgagcg gccgcgcac gtctcaaccg tgccgtgca tgaaatcctg gccggtttgt 2220
ctgatgccaa gtcgcggcc cttggcgccg gcttggcgcc tgaagaaacc gagcgccgcc 2280
gtctaaaaag gtgatgtgta tttgagtaa acagcttgcg tcatgcggtc gctgcgtata 2340
tgatgcgatg agtaataaaa caaatacgca aggggaacgc atgaagggtta tcgctgtact 2400
taaccagaaa ggccgggtcag gcaagacgac catcgcaacc catctagccc gcgcccgtga 2460
actcgccgg gccgatgttc tgttagtcga ttccgatccc cagggcagtg cccgctgctg 2520
ggcgcccggt cgggaagatc aaccgctaac cgttgtcgcc atcgaccgcc cgacgattga 2580
ccgcgacgtg aaggccatcg gccggcgcg cttcgtagt atcgacggag cgcccaggc 2640
ggcggaactg gctgtgtccg cgatcaaggc agccgacttc gtgctgattc cgggtgcagg 2700
aagcccttac gacatatggg ccaccgccta cctggtggag ctggttaagc agcgattga 2760
ggtcacggat ggaaggctac aagcgccctt tgtcgtgtcg cgggcgatca aaggcacgg 2820
catcgcggtt gaggttgccg aggcgctggc cgggtacgag ctgcccattc ttgagtcagg 2880
tatcacgcag cgcgtgagct accaggcac tgccgcggcc ggcacaaccg ttcttgaatc 2940
agaaccggag ggcgacgctg cccgcgaggt ccaggcgctg gccgtgaaa ttaaatcaaa 3000
actcatttga gttaatgagg taaagagaaa atgagcaaaa gcacaaacac gctaagtgcc 3060
ggccgtccga gcgcacgcag gacgaaggct gcaacgttgg ccagcctggc agacacgcca 3120
gccatgaagc ggggtcaactt tcagttgccg gcggaggatc acaccaagct gaagatgtac 3180
gcggtacgcc aaggcaagac cattaccgag ctgctatctg aatacatcgc gcagctacca 3240
gagtaaatga gcaaatgaat aaatgagtag atgaatttta gcggtataag gaggcggcat 3300
ggaaaatcaa gaaccaaccg gaccgacgc cgtggaatgc cccatgtgtg gaggaacgg 3360
cggttggcca ggcgtaagcg gctgggttgt ctgcggccc tgcaatggca ctggaacccc 3420
caagcccgag gaatcgcggt gacggtcgca aaccatccgg cccggtacaa atcggcggc 3480
cgctgggtga tgacctggtg gagaagttga aggcgcgca ggccgcccag cggcaacgca 3540
tcgaggcaga agcagcccc ggtgaatcgt ggcaagcgcc cgctgatcga atccgcaag 3600
aatcccgcca acccgcgga cccggtgcgc cgtcgattag gaagccgccc aaggcgacg 3660
agcaaccaga ttttttcgtt ccgatgctct atgacgtggg caccgcgat agtcgcagca 3720
tcattggacgt ggccgttttc cgtctgtcga agcgtgaccg acgagctggc gaggtgatcc 3780
gctacgagct tccagacggg cacgtagagg tttccgcagg gccggccggc atggccagtg 3840

```

tgtgggatta	cgacctggta	ctgatggcgg	tttcccatct	aaccgaatcc	atgaaccgat	3900
accgggaagg	gaaggggagac	aagcccggcc	gcgtgttccg	tccacacgtt	gcggacgtac	3960
tcaagttctg	ccggcgagcc	gatggcggaa	agcagaaaga	cgacctggta	gaaacctgca	4020
ttcggttaaa	caccacgcac	gttgccatgc	agcgtacgaa	gaaggccaag	aacggccgcc	4080
tggtagcgg	atccgaggg	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	4140
ccgggcccgc	ggagtacatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	4200
aaggcaagaa	cccgagcgtg	ctgacgggtc	accccgatta	ctttttgata	gatcccggca	4260
tccgcccgtt	tctctaccgc	ctggcacgcc	gcgcgcgagg	caaggcagaa	gccagatgg	4320
tgttcaagac	gatctacgaa	cgcagtggca	gcgcgcggaga	gttcaagaag	ttctgtttca	4380
ccgtgcgcaa	gctgatcggg	tcaaatgacc	tgccggagta	cgatttgaag	gaggaggcgg	4440
ggcaggctgg	cccgatccta	gtcatgcgct	accgcaacct	gatcgagggc	gaagcatccg	4500
ccggttccta	atgtacggag	cagatgctag	ggcaaatg	cctagcaggg	gaaaaaggtc	4560
gaaaaggctc	ctttcctgtg	gatagcacgt	acattgggaa	cccaaagccg	tacattggga	4620
accggaaccc	gtacattggg	aacccaaagc	cgtagattgg	gaaccgggtc	cacatgtaag	4680
tgactgatata	aaaagagaaa	aaaggcgatt	tttccgcta	aaactcttta	aaacttatta	4740
aaactcttaa	aacccgcctg	gcctgtgcat	aaactgtctg	ccagcgcaca	gccgaagagc	4800
tgcaaaaagg	gcctacccct	cggtcgctgc	gctccctacg	ccccgcgct	tcgcgtcggc	4860
ctatcgccgc	cgctggccgc	tcaaaaatgg	ctggcctacg	gccaggcaat	ctaccagggc	4920
gcggacaagc	cgcgccgctg	ccactcgacc	gccggcgccc	acatcaaggc	accctgcctc	4980
gcgcgtttcg	gtgatgacgg	tgaaaacctc	tgacacatgc	agctcccggg	gacggtcaca	5040
gcttgtctgt	aagcggatgc	cgggagcaga	caagcccgtc	agggcgcgct	agcgggtgtt	5100
ggcgggtgtc	ggggcgagc	catgacccag	tcacgtagcg	atagcggagt	gtatactggc	5160
ttactatgc	ggcatcagag	cagattgtac	tgagagtga	ccatatgcgg	tgtgaaatac	5220
cgcacagatg	cgtaaggaga	aaataccgca	tcaggcgctc	ttccgcttcc	tcgctcactg	5280
actcgctgcg	ctcggtcggt	cggctgcggc	gagcggatc	agctcactca	aaggcggtaa	5340
tacggttatc	cacagaatca	ggggataaac	caggaaagaa	catgtgagca	aaaggccagc	5400
aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcgtt	tttccatagg	ctccgcccc	5460
ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagaggtg	gcgaaacccg	acaggactat	5520
aaagatacca	ggcggtttccc	cctggaagct	ccctcgtgcg	ctctcctggt	ccgaccctgc	5580
cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	cgtggcgctt	tctcatagct	5640
cacgctgtag	gtatctcagt	tcgggtgtag	tcgttcgctc	caagctgggc	tgtgtgcacg	5700
aaccccccg	tcagcccgac	cgctgcgcct	tatccggtaa	ctatcgtctt	gagtccaacc	5760
cggtaaagaca	cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	5820
ggtatgtagg	cggtgctaca	gagttcttga	agtgggtggc	taactacggc	tacactagaa	5880
ggacagtatt	tggtatctgc	gctctgctga	agccagttac	cttcggaaaa	agagttggta	5940
gctcttgatc	cgccaaacaa	accaccgctg	gtagcgggtg	tttttttgg	tgcaagcagc	6000
agattacgcg	cagaaaaaaa	ggatctcaag	aagatccctt	gatcttttct	acgggggtctg	6060
acgctcagtg	gaacgaaaac	tcacgttaag	ggattttgg	catgcattct	aggtactaaa	6120
acaattcatc	cagtaaaata	taatatttta	ttttctccca	atcaggcttg	atccccagta	6180
agtcaaaaaa	tagctcgaca	tactgttctt	ccccgatata	ctccctgata	gaccgcagcg	6240
agaaaggcaat	gtcataccac	ttgtccgccc	tgccgcttct	cccaagatca	ataaagccac	6300
ttactttggc	atctttcaca	aagatgttgc	tgtctcccag	gtcgccgctg	gaaaagacaa	6360
gttctctctc	gggcttttcc	gtctttaaaa	aatcatacag	ctcgccggga	tctttaaatg	6420
gagtgtcttc	ttcccagttt	tcgcaatcca	catcgccag	atcgttattc	agtaagtaat	6480
ccaatccggc	taagcggctg	tctaagctat	tcgtatagg	acaatccgat	atgtcgattg	6540
agtgaagagc	cctgatgcac	tccgcataca	gctcgataat	cttttcaggg	ctttgttcat	6600
cttcatactc	ttccgagcaa	aggacgccat	cggcctcact	catgagcaga	ttgtctccagc	6660
catcatgccc	ttcaaagtgc	aggacctttg	gaacaggcag	ctttccttcc	agccatagca	6720
tcattgtcctt	ttcccgttcc	acatcatagg	tggtcccttt	ataccggctg	tcgctcattt	6780
ttaaatatag	gttttccattt	tctcccacca	gcttatatac	cttagcagga	gacattccct	6840
ccgtatcttt	tacgcagcgg	tatttttctga	tcagtttttt	caattccgg	gatattctca	6900
ttttagccat	ttattatttc	cttccctctt	tctacagtat	ttaaagatac	cccaagaagc	6960
taattataac	aagacgaact	ccaattcact	gttcccttga	ttctaaaaac	ttaaatacca	7020
gaaaacagct	ttttcaaagt	tgttttcaaa	gttggcgat	aacatagtat	cgacggagcc	7080
gattttgaaa	ccgcgggtgat	cacaggcagc	aacgctctgt	catcgttaca	atcaacatgc	7140
taccctccgc	gagatcatcc	gtgtttcaaa	cccgccagct	tagttgccgt	tcttccgaat	7200
agcatcggtg	acatgagcaa	agtctgccgc	cttacaacgg	ctctcccgt	gacgcgctcc	7260
cggactgatg	ggctgcctgt	atcgagtgg	gattttgtgc	cgagctgccg	gtcggggagc	7320
tgttggctgg	ctgggtggcag	gatataattg	ggtgtaaaac	aattgacgt	tagacaactt	7380
aataacacat	tgcggaagctt	tttaattgtac	tgaattaaac	ccgaattaat	tcgggggagc	7440
tggatttttag	tactggattt	tgggttttag	aattagaaat	tttattgata	gaagtatttt	7500
acaaatacaa	atacatacta	agggtttctt	atatgctcaa	cacatgagcg	aaacctata	7560
ggaaccctaa	ttcccttatac	tgggaactac	tcacacatta	ttatggagaa	actcgagctt	7620
gtcgatcgac	agatccggtc	ggcatctact	ctatttcttt	gcccctggac	gagtgctggg	7680
gcgtcggttt	ccactatcgg	cgagtacttc	tacacagcca	tcgggtccaga	cggcccgctc	7740
tctgcggcgg	atttgtgtac	gcccgcagct	cccggtccg	gatcggaaga	ttgcgtcgca	7800
tcgaccctgc	gcccgaagctg	catcatcgaa	attgccgtca	accaagctct	gatagagttg	7860

-60-

```

gtcaagacca atgcgaggca tatacgcccg gagtcgtggc gatcctgcaa gctccggatg 7920
cctccgctcg aagtagcgcg tctgctgctc catacaagcc aaccacggcc tccagaagaa 7980
gatggttggcg acctcggtatt gggaatcccc gaacatcgcc tcgctccagt caatgaccgc 8040
tggttatgctg ccattgtccg tcaggacatt gttggagccg aaatccgctg gcacgagggtg 8100
ccggacttcg gggcagtcct cggcccaaag catcagctca tcgagagcct gcgcgacgga 8160
cgcactgacg gtgtcgtcca tcacagtttg ccagtgtatc acatggggat cagcaatcgc 8220
gcatatgaaa tcacgccatg tagtgatttg accgattcct tgcggtccga atgggcccga 8280
cccgtcgtc tggctaagat cggccgcagc gatcgcatcc atagcctccg cgaccgggtg 8340
tagaacagcg ggcagttcgg tttcaggcag gtcttgcaac gtgacaccct gtgcacggcg 8400
ggagatgcaa taggtcaggc tctcgctaaa ctccccaatg tcaagcactt ccggaatcgg 8460
gagcgcggcc gatgcaaagt gccgataaac ataacgatct ttgtagaaac catcggcgca 8520
gctattttacc cgcaggacat atccacgccc tccatcacatg aagctgaaag caccgagattc 8580
ttcgccctcc gagagctgca tcaggctcga gacgctgtcg aactttttcga tcagaaactt 8640
ctcgacagac gtgcggtgga gttcaggctt tttcatatct cattgcccc ccggtatctgc 8700
gaaagctcga gagagataga tttgtagaga gagactgggtg atttcagcgt gtcctctcca 8760
aatgaaatga acttccttat atagaggaa gtcttgcgaa ggatagtggg attgtgcgtc 8820
atcccttacg tcagtgga taccacatca atccacttg atccacttg tttgaagacg tgggtggaaac 8880
gtcttctttt tccacgatgc tctcgtggg tgggggtcca tctttgggac cactgtcggc 8940
agaggcatct tgaacgatag cctttccttt atcgcaatga tggcatttgt aggtgccacc 9000
ttccttttct actgtccttt tgatgaagt acagatagct gggcaatgga atccgaggag 9060
gtttcccgat attacccttt gttgaaaagt ctcaatagcc ctttggcttt ctgagactgt 9120
atctttgata ttcttgagat gtcgtgctcc accatgttat cacatcaatc 9180
cacttgcttt gaagacgtgg ttggaacgtc ttctttttcc acgatgctcc tcgtgggtgg 9240
gggtccatct ttggggaccac tgtcggcaga ggcattctga acgatagcct ttcctttatc 9300
gcaatgatgg catttgtagg tgccaccttc cttttctact gtctttttga tgaagtga 9360
gatagctggg caatggaatc cgaggaggtt tcccgatatt accctttgtt gaaaagtctc 9420
aatagccctt tggctctctg agactgtatc tttgatattc ttggagtaga cgagagtgtc 9480
gtgctccacc atgttggtgca gctgctctag ccaatacgca aaccgcctct cccgcgcgt 9540
tggccgatcc attaatgcag ctggcacgac aggtttcccg actggaaagc gggcagtgag 9600
cgcaacgcaa ttaatgtgag ttagctcact cattaggcac cccaggcttt acactttatg 9660
cttcggctc gtatgttggt tggaaattgt agcggataac aatttcacac aggaaacagc 9720
tatgaccatg attacgaatt cgagctcggg acccggggat cctctagact gaaggcggga 9780
aacgacaatc tgatcatgag cggagaatta agggagtcac gttatgacct ccgccgatga 9840
cgccgggaaa gccgttttac gtttggaact gacagaaccg caacgttgaa ggagccactc 9900
agccgcgggt ttctggagtt taatgagcta agcacatacg tcagaaacca ttattgcgcg 9960
ttcaaaagtc ccctaaggtc actatcagct agcaaatatt tcttgtaaaa aatgtctcac 10020
tgacgttcca taaattcccc tcggtatcca attagagtct catattcact ctcaatccaa 10080
ataatctgca ccggtatctg agaatcgaat tccgcgggcc gc 10122

```

<210> 98

<211> 621

<212> DNA

<213> Artificial Sequence

<220>

<223> N. tabacum rDNA intergenic spacer (IGS) sequence

<300>

<308> Genbank #Y08422

<309> 1997-10-31

<400> 98

```

gtgctagcca atgtttaaca agatgtcaag cacaatgaat gttggtgggt ggtgggtcgtg 60
gctggcggtg gtggaaaatt gcggtggttc gagcggtagt gatcggcgat ggttgggtgt 120
tgcagcggtg ttgatatcg gaatcactta tgggtggttg cacaatggag gtgcgtcatg 180
gttattgggtg gttgggtcatc tatatatatt tataataata ttaagtattt tacctatttt 240
ttacatatatt tttattaaat ttatgcattg tttgtatttt taaatagttt ttatcgta 300
tgttttataa aatattttat tttttatgt gttatattat tacttgatgt attggaaatt 360
ttctccattg ttctttctat atttataata attttcttat tttttttgt tttattatgt 420
attttttcgt tttataataa atattttatta aaaaaaatat tatttttgta aaatatatca 480
tttacaatgt ttaaaagtca tttgtgaata tattagctaa gttgtacttc tttttgtgca 540
tttgggtgtg tacatgtcta ttatgattct ctggccaaaa catgtctact cctgtcactt 600
gggttttttt ttttaagaca t 621

```

<210> 99

<211> 25

<212> DNA

-61-

<213> Artificial Sequence

<220>

<223> NTIGS-F1 Primer

<400> 99

gtgctagcca atgtttaaca agatg 25

<210> 100

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> NTIGS-R1 Primer

<400> 100

atgtcttaaa aaaaaaaacc caagtgc 28

<210> 101

<211> 233

<212> DNA

<213> Mus Musculus

<300>

<308> Genbank #V00846

<309> 1989-07-06

<400> 101

gacctggaat atggcgagaa aactgaaaat cacggaaaat gagaaataca cacttttagga 60
 cgtgaaatat ggcgaggaaa actgaaaaag gtggaaaatt tagaaatgtc cactgttagga 120
 cgtggaatat ggcaagaaaa ctgaaaatca tggaaaatga gaaacatcca cttgacgact 180
 tgaaaaatga cgaaatcact aaaaaacgtg aaaaatgaga aatgcacact gaa 233

<210> 102

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> MSAT-F1 Primer

<400> 102

aataccgcgg agcttgacc tggaatatcg c 31

<210> 103

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> MSAT-R1 Primer

<400> 103

ataaccgcgg agtccttcag tgtgcat 27

<210> 104

<211> 277

<212> DNA

<213> Artificial Sequence

<220>

<223> Nopaline Synthase Promoter Sequence

<300>

<308> Genbank #U09365

<309> 1997-10-17

-62-

```

<400> 104
gagctcgaat ttccccgatac gttcaaacaat ttggcaataa agttttcttaa gattgaatcc 60
tggtgcccggg cttgcgatga ttatcatata atttctgttg aattacgtta agcatgtaat 120
aattaacatg taatgcatga cgttatttat gagatggggtt tttatgatta gaggccccgca 180
attatacatt taatacgcga tagaaaacaa aatatagcgc gcaaactagg ataaattatc 240
gcgcgcgggtg tcatctatgt tactagatcg ggaattc 277

<210> 105
<211> 1812
<212> DNA
<213> Escherichia coli

<220>
<221> CDS
<222> (1) ... (1812)
<223> Beta-Glucuronidase

<300>
<308> Genbank #S69414
<309> 1994-09-23

<400> 105
atg tta cgt cct gta gaa acc cca acc cgt gaa atc aaa aaa ctc gac 48
Met Leu Arg Pro Val Glu Thr Pro Thr Arg Glu Ile Lys Lys Leu Asp
1 5 10 15

ggc ctg tgg gca ttc agt ctg gat cgc gaa aac tgt gga att gat cag 96
Gly Leu Trp Ala Phe Ser Leu Asp Arg Glu Asn Cys Gly Ile Asp Gln
20 25 30

cgt tgg tgg gaa agc gcg tta caa gaa agc cgg gca att gct gtg cca 144
Arg Trp Trp Glu Ser Ala Leu Gln Glu Ser Arg Ala Ile Ala Val Pro
35 40 45

ggc agt ttt aac gat cag ttc gcc gat gca gat att cgt aat tat gcg 192
Gly Ser Phe Asn Asp Gln Phe Ala Asp Ala Asp Ile Arg Asn Tyr Ala
50 55 60

ggc aac gtc tgg tat cag cgc gaa gtc ttt ata ccg aaa ggt tgg gca 240
Gly Asn Val Trp Tyr Gln Arg Glu Val Phe Ile Pro Lys Gly Trp Ala
65 70 75 80

ggc cag cgt atc gtg ctg cgt ttc gat gcg gtc act cat tac ggc aaa 288
Gly Gln Arg Ile Val Leu Arg Phe Asp Ala Val Thr His Tyr Gly Lys
85 90 95

gtg tgg gtc aat aat cag gaa gtg atg gag cat cag ggc ggc tat acg 336
Val Trp Val Asn Asn Gln Glu Val Met Glu His Gln Gly Gly Tyr Thr
100 105 110

cca ttt gaa gcc gat gtc acg ccg tat gtt att gcc ggg aaa agt gta 384
Pro Phe Glu Ala Asp Val Thr Pro Tyr Val Ile Ala Gly Lys Ser Val
115 120 125

cgt atc acc gtt tgt gtg aac aac gaa ctg aac tgg cag act atc ccg 432
Arg Ile Thr Val Cys Val Asn Asn Glu Leu Asn Trp Gln Thr Ile Pro
130 135 140

ccg gga atg gtg att acc gac gaa aac ggc aag aaa aag cag tct tac 480
Pro Gly Met Val Ile Thr Asp Glu Asn Gly Lys Lys Lys Gln Ser Tyr
145 150 155 160

ttc cat gat ttc ttt aac tat gcc gga atc cat cgc agc gta atg ctc 528
Phe His Asp Phe Phe Asn Tyr Ala Gly Ile His Arg Ser Val Met Leu
165 170 175

tac acc acg ccg aac acc tgg gtg gac gat atc acc gtg gtg acg cat 576

```

-63-

Tyr	Thr	Thr	Pro	Asn	Thr	Trp	Val	Asp	Asp	Ile	Thr	Val	Val	Thr	His	
			180					185					190			
gtc	gcg	caa	gac	tgt	aac	cac	gcg	tct	gtt	gac	tgg	cag	gtg	gtg	gcc	624
Val	Ala	Gln	Asp	Cys	Asn	His	Ala	Ser	Val	Asp	Trp	Gln	Val	Val	Ala	
		195					200					205				
aat	ggt	gat	gtc	agc	gtt	gaa	ctg	cgt	gat	gcg	gat	caa	cag	gtg	gtt	672
Asn	Gly	Asp	Val	Ser	Val	Glu	Leu	Arg	Asp	Ala	Asp	Gln	Gln	Val	Val	
	210					215					220					
gca	act	gga	caa	ggc	act	agc	ggg	act	ttg	caa	gtg	gtg	aat	ccg	cac	720
Ala	Thr	Gly	Gln	Gly	Thr	Ser	Gly	Thr	Leu	Gln	Val	Val	Asn	Pro	His	
225					230					235					240	
ctc	tgg	caa	ccg	ggt	gaa	ggt	tat	ctc	tat	gaa	ctg	tgc	gtc	aca	gcc	768
Leu	Trp	Gln	Pro	Gly	Glu	Gly	Tyr	Leu	Tyr	Glu	Leu	Cys	Val	Thr	Ala	
				245					250					255		
aaa	agc	cag	aca	gag	tgt	gat	atc	tac	ccg	ctt	cgc	gtc	ggc	atc	cgg	816
Lys	Ser	Gln	Thr	Glu	Cys	Asp	Ile	Tyr	Pro	Leu	Arg	Val	Gly	Ile	Arg	
			260					265					270			
tca	gtg	gca	gtg	aag	ggc	gaa	cag	ttc	ctg	att	aac	cac	aaa	ccg	ttc	864
Ser	Val	Ala	Val	Lys	Gly	Glu	Gln	Phe	Leu	Ile	Asn	His	Lys	Pro	Phe	
		275					280					285				
tac	ttt	act	ggc	ttt	ggt	cgt	cat	gaa	gat	gcg	gac	ttg	cgt	ggc	aaa	912
Tyr	Phe	Thr	Gly	Phe	Gly	Arg	His	Glu	Asp	Ala	Asp	Leu	Arg	Gly	Lys	
	290					295					300					
gga	ttc	gat	aac	gtg	ctg	atg	gtg	cac	gac	cac	gca	tta	atg	gac	tgg	960
Gly	Phe	Asp	Asn	Val	Leu	Met	Val	His	Asp	His	Ala	Leu	Met	Asp	Trp	
305					310					315					320	
att	ggg	gcc	aac	tcc	tac	cgt	acc	tcg	cat	tac	cct	tac	gct	gaa	gag	1008
Ile	Gly	Ala	Asn	Ser	Tyr	Arg	Thr	Ser	His	Tyr	Pro	Tyr	Ala	Glu	Glu	
				325					330					335		
atg	ctc	gac	tgg	gca	gat	gaa	cat	ggc	atc	gtg	gtg	att	gat	gaa	act	1056
Met	Leu	Asp	Trp	Ala	Asp	Glu	His	Gly	Ile	Val	Val	Ile	Asp	Glu	Thr	
			340					345					350			
gct	gct	gtc	ggc	ttt	aac	ctc	tct	tta	ggc	att	ggt	ttc	gaa	gcg	ggc	1104
Ala	Ala	Val	Gly	Phe	Asn	Leu	Ser	Leu	Gly	Ile	Gly	Phe	Glu	Ala	Gly	
		355					360					365				
aac	aag	ccg	aaa	gaa	ctg	tac	agc	gaa	gag	gca	gtc	aac	ggg	gaa	act	1152
Asn	Lys	Pro	Lys	Glu	Leu	Tyr	Ser	Glu	Glu	Ala	Val	Asn	Gly	Glu	Thr	
	370					375					380					
cag	caa	gcg	cac	tta	cag	gcg	att	aaa	gag	ctg	ata	gcg	cgt	gac	aaa	1200
Gln	Gln	Ala	His	Leu	Gln	Ala	Ile	Lys	Glu	Leu	Ile	Ala	Arg	Asp	Lys	
385					390					395					400	
aac	cac	cca	agc	gtg	gtg	atg	tgg	agt	att	gcc	aac	gaa	ccg	gat	acc	1248
Asn	His	Pro	Ser	Val	Val	Met	Trp	Ser	Ile	Ala	Asn	Glu	Pro	Asp	Thr	
				405					410					415		
cgt	ccg	caa	ggt	gca	cgg	gaa	tat	ttc	gcg	cca	ctg	gcg	gaa	gca	acg	1296
Arg	Pro	Gln	Gly	Ala	Arg	Glu	Tyr	Phe	Ala	Pro	Leu	Ala	Glu	Ala	Thr	
			420					425					430			
cgt	aaa	ctc	gac	ccg	acg	cgt	ccg	atc	acc	tgc	gtc	aat	gta	atg	ttc	1344
Arg	Lys	Leu	Asp	Pro	Thr	Arg	Pro	Ile	Thr	Cys	Val	Asn	Val	Met	Phe	
		435					440					445				

-64-

tgc gac gct cac acc gat acc atc agc gat ctc ttt gat gtg ctg tgc	1392
Cys Asp Ala His Thr Asp Thr Ile Ser Asp Leu Phe Asp Val Leu Cys	
450 455 460	
ctg aac cgt tat tac gga tgg tat gtc caa agc ggc gat ttg gaa acg	1440
Leu Asn Arg Tyr Tyr Gly Trp Tyr Val Gln Ser Gly Asp Leu Glu Thr	
465 470 475 480	
gca gag aag gta ctg gaa aaa gaa ctt ctg gcc tgg cag gag aaa ctg	1488
Ala Glu Lys Val Leu Glu Lys Glu Leu Leu Ala Trp Gln Glu Lys Leu	
485 490 495	
cat cag ccg att atc atc acc gaa tac ggc gtg gat acg tta gcc ggg	1536
His Gln Pro Ile Ile Ile Thr Glu Tyr Gly Val Asp Thr Leu Ala Gly	
500 505 510	
ctg cac tca atg tac acc gac atg tgg agt gaa gag tat cag tgt gca	1584
Leu His Ser Met Tyr Thr Asp Met Trp Ser Glu Glu Tyr Gln Cys Ala	
515 520 525	
tgg ctg gat atg tat cac cgc gtc ttt gat cgc gtc agc gcc gtc gtc	1632
Trp Leu Asp Met Tyr His Arg Val Phe Asp Arg Val Ser Ala Val Val	
530 535 540	
ggt gaa cag gta tgg aat ttc gcc gat ttt gcg acc tcg caa ggc ata	1680
Gly Glu Gln Val Trp Asn Phe Ala Asp Phe Ala Thr Ser Gln Gly Ile	
545 550 555 560	
ttg cgc gtt ggc ggt aac aag aaa ggg atc ttc act cgc gac cgc aaa	1728
Leu Arg Val Gly Asn Lys Lys Gly Ile Phe Thr Arg Asp Arg Lys	
565 570 575	
ccg aag tcg gcg gct ttt ctg ctg caa aaa cgc tgg act ggc atg aac	1776
Pro Lys Ser Ala Ala Phe Leu Leu Gln Lys Arg Trp Thr Gly Met Asn	
580 585 590	
ttc ggt gaa aaa ccg cag cag gga ggc aaa caa tga	1812
Phe Gly Glu Lys Pro Gln Gln Gly Gly Lys Gln *	
595 600	

<210> 106

<211> 603

<212> PRT

<213> Escherichia coli

<300>

<308> Genbank #S69414

<309> 1994-09-23

<400> 106

Met Leu Arg Pro Val Glu Thr Pro Thr Arg Glu Ile Lys Lys Leu Asp	
1 5 10 15	
Gly Leu Trp Ala Phe Ser Leu Asp Arg Glu Asn Cys Gly Ile Asp Gln	
20 25 30	
Arg Trp Trp Glu Ser Ala Leu Gln Glu Ser Arg Ala Ile Ala Val Pro	
35 40 45	
Gly Ser Phe Asn Asp Gln Phe Ala Asp Ala Asp Ile Arg Asn Tyr Ala	
50 55 60	
Gly Asn Val Trp Tyr Gln Arg Glu Val Phe Ile Pro Lys Gly Trp Ala	
65 70 75 80	
Gly Gln Arg Ile Val Leu Arg Phe Asp Ala Val Thr His Tyr Gly Lys	
85 90 95	
Val Trp Val Asn Asn Gln Glu Val Met Glu His Gln Gly Gly Tyr Thr	
100 105 110	
Pro Phe Glu Ala Asp Val Thr Pro Tyr Val Ile Ala Gly Lys Ser Val	
115 120 125	

-65-

Arg	Ile	Thr	Val	Cys	Val	Asn	Asn	Glu	Leu	Asn	Trp	Gln	Thr	Ile	Pro
	130					135					140				
Pro	Gly	Met	Val	Ile	Thr	Asp	Glu	Asn	Gly	Lys	Lys	Lys	Gln	Ser	Tyr
145				150						155					160
Phe	His	Asp	Phe	Phe	Asn	Tyr	Ala	Gly	Ile	His	Arg	Ser	Val	Met	Leu
				165					170					175	
Tyr	Thr	Thr	Pro	Asn	Thr	Trp	Val	Asp	Asp	Ile	Thr	Val	Val	Thr	His
			180					185					190		
Val	Ala	Gln	Asp	Cys	Asn	His	Ala	Ser	Val	Asp	Trp	Gln	Val	Val	Ala
	195						200					205			
Asn	Gly	Asp	Val	Ser	Val	Glu	Leu	Arg	Asp	Ala	Asp	Gln	Gln	Val	Val
	210					215					220				
Ala	Thr	Gly	Gln	Gly	Thr	Ser	Gly	Thr	Leu	Gln	Val	Val	Asn	Pro	His
225					230					235					240
Leu	Trp	Gln	Pro	Gly	Glu	Gly	Tyr	Leu	Tyr	Glu	Leu	Cys	Val	Thr	Ala
				245					250					255	
Lys	Ser	Gln	Thr	Glu	Cys	Asp	Ile	Tyr	Pro	Leu	Arg	Val	Gly	Ile	Arg
			260					265					270		
Ser	Val	Ala	Val	Lys	Gly	Glu	Gln	Phe	Leu	Ile	Asn	His	Lys	Pro	Phe
	275						280					285			
Tyr	Phe	Thr	Gly	Phe	Gly	Arg	His	Glu	Asp	Ala	Asp	Leu	Arg	Gly	Lys
	290					295					300				
Gly	Phe	Asp	Asn	Val	Leu	Met	Val	His	Asp	His	Ala	Leu	Met	Asp	Trp
305					310					315					320
Ile	Gly	Ala	Asn	Ser	Tyr	Arg	Thr	Ser	His	Tyr	Pro	Tyr	Ala	Glu	Glu
				325					330					335	
Met	Leu	Asp	Trp	Ala	Asp	Glu	His	Gly	Ile	Val	Val	Ile	Asp	Glu	Thr
			340					345					350		
Ala	Ala	Val	Gly	Phe	Asn	Leu	Ser	Leu	Gly	Ile	Gly	Phe	Glu	Ala	Gly
	355						360					365			
Asn	Lys	Pro	Lys	Glu	Leu	Tyr	Ser	Glu	Glu	Ala	Val	Asn	Gly	Glu	Thr
	370					375					380				
Gln	Gln	Ala	His	Leu	Gln	Ala	Ile	Lys	Glu	Leu	Ile	Ala	Arg	Asp	Lys
385					390					395					400
Asn	His	Pro	Ser	Val	Val	Met	Trp	Ser	Ile	Ala	Asn	Glu	Pro	Asp	Thr
				405					410					415	
Arg	Pro	Gln	Gly	Ala	Arg	Glu	Tyr	Phe	Ala	Pro	Leu	Ala	Glu	Ala	Thr
				420				425					430		
Arg	Lys	Leu	Asp	Pro	Thr	Arg	Pro	Ile	Thr	Cys	Val	Asn	Val	Met	Phe
	435						440					445			
Cys	Asp	Ala	His	Thr	Asp	Thr	Ile	Ser	Asp	Leu	Phe	Asp	Val	Leu	Cys
	450				455					460					
Leu	Asn	Arg	Tyr	Tyr	Gly	Trp	Tyr	Val	Gln	Ser	Gly	Asp	Leu	Glu	Thr
465					470					475					480
Ala	Glu	Lys	Val	Leu	Glu	Lys	Glu	Leu	Leu	Ala	Trp	Gln	Glu	Lys	Leu
				485					490					495	
His	Gln	Pro	Ile	Ile	Ile	Thr	Glu	Tyr	Gly	Val	Asp	Thr	Leu	Ala	Gly
			500					505					510		
Leu	His	Ser	Met	Tyr	Thr	Asp	Met	Trp	Ser	Glu	Glu	Tyr	Gln	Cys	Ala
	515						520					525			
Trp	Leu	Asp	Met	Tyr	His	Arg	Val	Phe	Asp	Arg	Val	Ser	Ala	Val	Val
	530					535					540				
Gly	Glu	Gln	Val	Trp	Asn	Phe	Ala	Asp	Phe	Ala	Thr	Ser	Gln	Gly	Ile
545					550					555					560
Leu	Arg	Val	Gly	Gly	Asn	Lys	Lys	Gly	Ile	Phe	Thr	Arg	Asp	Arg	Lys
				565					570					575	
Pro	Lys	Ser	Ala	Ala	Phe	Leu	Leu	Gln	Lys	Arg	Trp	Thr	Gly	Met	Asn
			580					585					590		
Phe	Gly	Glu	Lys	Pro	Gln	Gln	Gly	Gly	Lys	Gln					
	595						600								

<210> 107

<211> 277

<212> DNA

<213> Artificial Sequence

-66-

<220>
 <223> Nopaline Synthase Terminator Sequence

<300>
 <308> U09365
 <309> 1995-10-17

<400> 107
 gagctcgaat ttccccgatac gttcaaacat ttggcaataa agtttcttaa gattgaatcc 60
 tgttgccggt cttgcgatga ttatcatata atttctgttg aattacgtta agcatgtaat 120
 aattaacatg taatgcatga cgttatttat gagatgggtt tttatgatta gagtcccgca 180
 attatacatg taatacgcga tagaaaacaa aatatagcgc gcaaactagg ataaattatc 240
 gcgcgcgggtg tcattctatgt tactagatcg ggaattc 277

<210> 108
 <211> 3451
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> HindIII Fragment containing the beta-glucuronidase coding sequence, the rDNA intergenic spacer, and the MastI sequence

<400> 108
 aagcttgacc tggaaatcgc cgagtaaact gaaaatcacg gaaaatgaga aatacacact 60
 ttaggacgtg aaatatggcg aggaaaactg aaaaagggtg aaaatttaga aatgtccact 120
 gtaggacgtg gaatatggca agaaaactga aaatcatgga aaatgagaaa catccacttg 180
 acgacttgaa aaatgacgaa atcactaaaa aacgtgaaaa atgagaaatg cacactgaag 240
 gactccgcgg gaattcgatt gtgctagcca atgtttaaca agatgtcaag cacaatgaat 300
 gttgggtggt ggtggtcggt gctggcggtg gtggaaaatt gcggtggttc gagcggttagt 360
 gatcgcgcat ggttgggtgt tgcagcggtg tttgatatcg gaatcactta tgggtggttgt 420
 cacaatggag gtgcgtcatg gttattgggt gttggtcatt tatatatatt tataataata 480
 ttaagtattt tacctatttt ttacatattt tttattaaat ttatgcattg tttgtatttt 540
 taaatagttt ttatcgctact tgtttttataa aatattttat tattttatgt gttatattat 600
 tacttgatgt attggaaatt ttctccattg ttttttctat atttataata attttcttat 660
 ttttttttgt tttattatgt attttttcgt tttataataa atatttatta aaaaaaatat 720
 tatttttgtt aaatatatca tttacaatgt ttaaaagtca tttgtgaata tattagctaa 780
 gttgtacttc tttttgtgca tttgggtgtg tacatgtcta ttatgattct ctggccaaaa 840
 catgtctact cctgtcactt ggggtttttt ttttaagaca taatcactag tgattatatac 900
 tagactgaag gcgggaaacg acaatctgat catgagcgga gaattaaggg agtcacgtta 960
 tgacccccgc cgatgacgcg ggacaagccg ttttacgttt ggaactgaca gaaccgcaac 1020
 gttgaaggag ccactcagcc gcgggtttct ggagtttaat gagctaagca catacgtcag 1080
 aaaccattat tgcgcgttca aaagtcgcct aaggtcacta tcagctagca aatatttctt 1140
 gtcaaaaatg ctccactgac gttccataaa ttcccctcgg tatccaatta gagtctcata 1200
 ttcactctca atccaaataa tctgcaccgg atctcgagat cgaattcccg cgccgcgcaa 1260
 ttcactagtg gatccccggg tacgggtcagt cccttatgtt acgtcctgta gaaaccccaa 1320
 cccgtgaaat caaaaaactc gacggcctgt gggcattcag tctggatcgc gaaaactgtg 1380
 gaattgagca gcgttgggtg gaaagcgcgt tacaagaaag ccgggcaatt gctgtgccag 1440
 gcagttttta cgatcagttc gccgatgcag atattcgtaa ttatgtgggc aacgtctggt 1500
 atcagcgcga agtctttata ccgaaagggt gggcaggcca gcgtatcgtg ctgcgtttcg 1560
 atgcggtcac tcattacggc aaagtgtggg tcaataatca ggaagtgatg gagcatcagg 1620
 gcggctatac gccatttgaa gccgatgtca cgccgtatgt tattgccggg aaaagtgtac 1680
 gtatcacagt ttgtgtgaac aacgaactga actggcagac tatcccgccg ggaatgggtga 1740
 ttaccgacga aacaggcaag aaaaagcagt cttacttcca tgatttcttt aactacgcgc 1800
 ggatccatcg cagcgtaatg ctctacacca cgccgaacac ctgggtggac gatatacccg 1860
 tgggtgacga tgcgcgcgaa gactgtaacc acgcgtctgt tgactggcag gtggtggcca 1920
 atggtgatgt cagcgttgaa ctgcgtgatg cggatcaaca ggtggttgca actggacaag 1980
 gcaccacggg gcatttgcaa gtggtgaaatc gcacactctg gcaaccgggt gaaggttatc 2040
 tctatgaact gtacgtcaca gccaaaagcc agacagagtg tgatatctac ccgctgcgcg 2100
 tcggcatccg gtcagtggca gtgaagggcg aacagttcct gatcaaccac aaaccgttct 2160
 actttactgg ctttggccgt catgaagatg cggatttcgc cggcaaaagg ttcgataacg 2220
 tgcgtatggt gcaagatcac gcattaatgg actggattgg ggccaactcc taccgtacct 2280
 cgcattaccc ttacgctgaa gagatgctcg actgggcaga atcgtgggtga 2340
 ttgatgaaac tgcagctgtc ggcttttaacc tctctttagg cattgggttc gaagcgggca 2400
 acaagccgaa agaactgtac agcgaagagg cagtcaacgg ggaaactcag caggcgcact 2460
 tacaggcgat taaagagctg atagcgcgtg acaaaaacca cccaagcgtg gtgatgtgga 2520

-67-

```

gtattgccaa cgaaccggat acccggtccgc aaggtgcacg ggaatatttc ggcgcactgg 2580
cggaagcaac gcgtaaactc gatccgacgc gtccgatcac ctgcgtcaat gtaatgttct 2640
gcgacgctca caccgatacc atcagcgatc tctttgatgt gctgtgcctg aaccgttatt 2700
acggttggtg tgtccaaagc ggcgatttgg aaacggcaga gaaggtactg gaaaaagaac 2760
ttctggcctg gcaggagaaa ctgcatcagc cgattatcat caccgaatac ggcgtggata 2820
cgtagccggg gctgcactca atgtacaccg acatgtggag tgaagagtat cagtgtgcat 2880
ggctggatat gtatcaccgc gtctttgatc gcgtcagcgc cgtcgtcggg gaacagggtat 2940
ggaatttcgc cgattttgcg acctcgcaag cggtggcggg aacaagaagg 3000
ggatcttcac ccgcgaccgc aaaccgaagt cggcggcctt tctgctgcaa aaacgctgga 3060
ctggcatgaa cttcgggtgaa aaaccgcagc agggaggcaa acaatgaatc aacaactctc 3120
ctggcgcaac atcgtcggct acagcctcgg gaattgcgta ccgagctcga atttccccga 3180
tcgttcaaac atttggcaat aaagtctctt aagattgaat cctgttgccg gtcttgcatg 3240
gattatcata taatttctgt tgaattacgt taagcatgta ataattaaca tgtaatgcat 3300
gacgttatatt atgagatggg ttttatgat tagagtcccg caattataca tttaatacgc 3360
gatagaaaac aaaatatagc gcgcaaaacta ggataaatta tcgcgcgcgg tgtcatctat 3420
gttactagat cgggaattcg atatcaagct t 3451

```

<210> 109

<211> 14627

<212> DNA

<213> Artificial Sequence

<220>

<223> pAg11a Plasmid

<400> 109

```

catgccaaac acagggttcc cctcgggatc aaagtacttt gatccaaccc ctccgctgct 60
atagtgcagt cggcttctga cgttcagtgc agccgtcttc tgaaaacgac atgtcgcaca 120
agtoctaagt tacgcgacag gctgccgccc tgcccttttc ctggcgcttt cttgtcgcgt 180
gttttagtgc cataaagtag aatacttgcc actagaaccg gagacattac gccatgaaca 240
agagcgccgc cgctggcctg ctgggctatg ccgcgtcag caccgacgac caggacttga 300
ccaaccaacg ggccgaactg cacgcggccg gctgcaccaa gctgttttcc gagaagatca 360
ccggcaccag gcgcgaccgc ccggagctgg ccaggatgct tgaccaccta cgccctggcg 420
acgttggtgac agtgaccagg ctagaccgcc tggcccgcag caccgcgcgac ctactggaca 480
ttgcccagcg catccaggag gccggcgccg gcctgcgtag cctggcagag ccgtgggccc 540
acaccaccac ccgggcccgc cgcatgggtg tgacogtggt cgcgggcatt gccgagttcg 600
agcgttccct aatcatcgac cgcacccgga gcgggcgcga ggccgccaag gcccgaggcg 660
tgaagtttgg ccccgccct accctcaccc cggcacagat cgcgcacgcc cgcgagctga 720
tcgaccagga aggcgcgacc gtgaaagagg cggctgcact gcttggcgtg catcgtcga 780
ccctgtaccg cgcacttgag cgcagcgagg aagtgcagcc caccgaggcc aggcggcgcg 840
gtgccttcgc tgaggacgca ttgaccgagg ccgacgccc ctggcgccgc gagaatgaac 900
gccaagagga acaagcatga aaccgcacca ggacggccag gacgaaccgt ttttcattac 960
cgaagagatc gaggcggaga tgatcgccgc cgggtacgtg ttogagccgc ccgcgcacgt 1020
ctcaaccgtg cggctgcatg aaatcctggc cggtttgtct gatgccaagc tggcggcctg 1080
gccggccagc ttggccgctg aagaaaaccg cgcgcgccgt ctaaaaaggt gatgtgtatt 1140
tgagtaaaac agcttgcgtc atgcccgtgc tgcgtatatg atgcgatgag taaataaaca 1200
aatacgcaag gggaacgcat gaaggttatc gctgtactta accagaaagg cgggtcaggc 1260
aagacgacca tcgcaaccca tctagcccgc gcctgcgaac tcgcgcgggg cgtgttctg 1320
ttagtcgatt ccgatcccca gggcagtgcc cgcgattggg cggcgtgctg ggaagatcaa 1380
ccgctaaccg ttgtcgcat cgaccgcccg acgattgacc cgcacgtgaa ggccatcgcc 1440
cggcgcgact tcgtagtgat cgacggagcg ccccaggcgg cggacttggc tgtgtccgcg 1500
atcaaggcag ccgacttcgt gctgattccg gtgagccaa gcccttacga catatgggcc 1560
accgcgcacc ttggtggagct ggttaagcag cgcattgagg tcacggatgg aaggctacaa 1620
gcggcctttg tcgtgtcgcg ggcgatcaaa ggcacgcgca tcggcggtga ggttgccgag 1680
gcgctggccg ggtacgagct gccatttctt gagtcccgtg tcacgcagcg cgtgagctac 1740
ccaggcactg ccgcgcgcgg cacaaccgtt cttgaatcag aaccgcaggg cgacgctgcc 1800
cgcgaggtcc aggcgctggc cgctgaaatt aaatcaaaac tcatttgagt taatgaggta 1860
aagagaaaaa gagcaaaaag acaaacacgc taagtgcggc ccgtccgagc gcacgcagca 1920
gcaaggctgc aacgttgcc agcctggcag acacgcagc catgaagcgg gtcaactttc 1980
agttgcgggc ggaggatcac accaagctga agatgtacgc ggtacgcca ggcaagacca 2040
ttaccgagct gctatctgaa tacatcgccg agctaccaga gtaaatgagc aaatgaataa 2100
atgagtagat gaatttttagc ggctaaagga ggcggcatgg aaaatcaaga acaaccaggc 2160
accgacgcgc tggaaatgcc catgtgtgga ggaacggcg gttggccagg cgtaaagcgc 2220
tgggttgtct gcggccctg caatggcact ggaaccccca agcccaggga atcgcgctga 2280
cggctcgcaa ccatccggcc cggtaaaaat cggcgcggcg ctgggtgatg acctggtgga 2340
gaagttgaag gccgcgcagg ccgcccagcg gcaacgcata gaggcagaag cacgcccgcg 2400
tgaatcgtgg caagcggccg ctgatcgaat ccgcaaagaa tcccggcaac cgccggcagc 2460

```

cgggtgcgccg	tgcgattagga	agccgcgccaa	gggcgacgag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tgcgagcatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	ggtgatccgc	tacgagcttc	cagacgggca	2640
cgtagagggt	tccgcagggc	cggccggcat	ggccagtggt	tgggattacg	acctgggtact	2700
gatggcggtt	tcccatctaa	ccgaatccat	gaaccgatac	cgggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgct	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaag	cagaaagacg	acctggtaga	aacctgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggccgcctg	gtgacggtat	ccgaggggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaacc	gggcggcccg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gacacagaaa	ggcaagaacc	cggacgtgct	3060
gacggttcac	cccgattact	ttttgatoga	tcccgcatc	ggccgttttc	tctaccgcct	3120
ggcacgccgc	gccgcaggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gccggagagt	tcaagaagtt	ctgtttcacc	gtgcgcaagc	tgatcgggtc	3240
aatgacctg	ccggagtacg	atattgaagga	ggaggcgggg	caggctggcc	cgatcctagt	3300
catgcgctac	cgcaacctga	tcgagggcga	agcatccgcc	ggttcctaata	gtacggagca	3360
gatgctaggg	caaattggcc	tagcagggga	aaaaggtcga	aaaggtctct	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaaaccgt	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	catgtaagtg	actgatataa	aagagaaaaa	3540
aggcgatatt	tccgcctaaa	actctttaaa	acttattaaa	actcttaaaa	cccgcctggc	3600
ctgtgcataa	ctgtctggcc	agcgacagc	cgaagagctg	caaaaagcgc	ctacccttcg	3660
gtcgctgcgc	tccctacgcc	ccgcgcgttc	cgctcgccct	atcgccggcc	ctggccgctc	3720
aaaaatggct	ggcctacggc	caggcaatct	accagggcgc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	cggcgcccac	atcaaggcac	cctgcctcgc	gcgtttcggg	gatgacgggtg	3840
aaaacctctg	acacatgcag	ctcccggaga	cggtcacagc	ttgtctgtaa	gcggatgccg	3900
ggagcagaca	agcccgtcag	ggcgcgctcag	cgggtgttgg	cgggtgtcgg	ggcgacagca	3960
tgaccagtc	acgtagcgat	agcggagtg	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcgggtg	tgaataaccg	cacagatgcg	taaggagaaa	4080
ataccgcac	aggcgctctt	ccgcttcctc	gctcactgac	tgcgtgcgct	cggtcggttcg	4140
gctgcggcga	gcggatcag	ctcactcaaa	ggcggtaata	cggttatcca	cagaatcagg	4200
ggataacgca	ggaaagaaca	tgtgagcaaa	agggcagcaa	aaggccagga	accgtaaaaa	4260
ggccgcgttg	ctggcgtttt	tccataggct	ccgccccct	gacgagcatc	acaaaaatcg	4320
acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	agataccagg	cgtttccccc	4380
tggaagctcc	ctcgtgcgct	ctcctgttcc	gaccctgccc	cttaccggat	acctgtccgc	4440
ctttctccct	tgggaagcg	tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	4500
gggtaggttc	agctggctcc	agctggcttc	tggtgcagaa	ccccccgttc	agcccgaccg	4560
ctgcgcctta	tccggttaact	atcgtcttga	gtccaaccgc	gtaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tgggtggccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4740
tctgctgaag	ccagttacct	tcggaaaaag	agttggtagc	tcttgatccg	gcaaacaaac	4800
caccgcgtgt	agcgggtggt	tttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	gctcagtgga	acgaaaactc	4920
acgttaaggg	attttggtca	tgcattctag	gtactaaaaa	aattcatcca	gtaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	5040
ctgttcttcc	cogatatcct	ccctgatcga	cggcagcgag	aaggcaatgt	cataccactt	5100
gtccgccttg	ccgcttctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgctg	tctcccaggt	cgccgtggga	aaagacaagt	tctcttccg	gcttttccgt	5220
ctttaaaaaa	tcatacagct	cgccgcggatc	tttaaatgga	gtgtcttctt	cccagttttc	5280
gcaatccaca	tccggccagat	cgttattcag	taagtaatcc	aattcggcta	agcggctgtc	5340
taagctattc	gtatagggac	aatccgatat	gtcgatggag	tgaaagagcc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggct	ttgttcactc	tcatactctt	ccgagcaaa	5460
gacgccatcg	gcctcactca	tgagcagatt	gctccagcca	tcatgccggt	caaagtgcag	5520
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5580
gatcatagtg	gtccctttat	accggctgtc	cgctattttt	aaatataggt	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgcagcggta	5700
tttttcgac	agttttttca	attccgggtga	tattctcatt	ttagccattt	attatttccct	5760
tcctcttttc	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tccttgcat	ctaaaacctt	aaataccaga	aaacagcttt	ttcaaagtgt	5880
ttttcaaaag	tggcgtataa	catagtatcg	acggagccga	ttttgaaacc	gcggtgatca	5940
caggcagcaa	cgctctgtca	tcgttacaat	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaacc	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaa	6060
tctgccgcct	tacaacggct	ctcccgtgta	cgccgtcccg	gactgatggg	ctgcctgtat	6120
cgagtgggtga	ttttgtgccg	agctgccggt	cggggagctg	ttggctggct	ggtggcagga	6180
tatatgttgg	tgtaaacaaa	ttgacgctta	gacaacttaa	taacacattg	cggacgtttt	6240
taatgtactg	aattaacgcc	gaattaattc	gggggatctg	gatttttagta	ctggattttg	6300
gttttaggaa	ttagaaattt	tattgataga	agtattttac	aaatacaaat	acatactaag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctatagg	aaccctaatt	cccttatctg	6420
ggaactactc	acacattatt	atggagaaac	tcgagtcaaa	tctcggtgac	gggcaggacc	6480

-69-

ggacggggcg	gtaccggcag	gctgaagtc	agctgccaga	aacccacgtc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgcctcgtgc	6600
atgcgcacgc	tcgggtcggt	gggcagcccg	atgcagcga	ccacgctcct	gaagccctgt	6660
gcctccagg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcctgt	ccgctgggtg	6720
cggggggaga	cgtacacggg	cgactcggcc	gtccagtcgt	aggcgttgcg	tgccttccag	6780
gggcccgcgt	aggcgatgcc	ggcgacctcg	ccgtccacct	cggcgacgag	ccaggatag	6840
cgctcccga	gacggacgag	gtcgtccgtc	cactcctgcg	gttccctgcg	ctcggtagcg	6900
aagttgaccg	tgcttgtctc	gatgtagtg	ttgacgatgg	tgcagaccgc	cggcatgtcc	6960
gcctcgggtg	cacggcggat	gtcggccggg	cgtcgttctg	ggctcatggg	agactcgaga	7020
gagatagatt	tgtagagaga	gactgggtgat	ttcagcgtgt	cctctccaaa	tgaaatgaac	7080
ttccttatat	agaggaaggt	cttgcgaagg	atagtgggat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaaacgt	cttctttttc	7200
cacgatgctc	ctcgtgggtg	gggggtccatc	tttgggacca	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttcctttat	cgcaatgatg	gcatttgtag	gtgccacctt	ccttttctac	7320
tgtccttttg	atgaagtgc	agatagctgg	gcaatggaat	ccgaggagggt	ttcccgata	7380
taccctttgt	tgaaaagtct	caatagccct	ttggctttct	gagactgtat	ctttgatatt	7440
cttggagtag	acgagagtg	cgtgctccac	catgttatca	catcaatcca	cttgctttga	7500
agacgtgggt	ggaacgtctt	ctttttccac	gatgctcctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcttgaac	gatagccttt	cctttatcgc	aatgatggca	7620
tttgtaggtg	ccaccttctt	tttctactgt	ccttttgatg	aagtgcagca	tagctgggca	7680
atggaatccg	aggaggtttc	ccgatattac	cctttgttga	aaagtctcaa	tagccctttg	7740
gtcttctgag	actgtatctt	tgtatattct	ggagtgcagc	agagtgtcgt	gtccaccat	7800
gttggcaagc	tgctctagcc	aatacgcaaa	ccgcctctcc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcacgcacg	gtttcccgac	tggaaagcgg	gcagtgcagc	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcaccc	caggctttac	actttatgct	tccggctcgt	7980
atgttgtgtg	gaattgtgag	cggataacaa	tttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agccttgact	agagggtcga	cggatatacag	acatgataag	atacattgat	8100
gagtttgagc	aaaccacaa	tagaatgcag	tgaaaaaaat	gctttatttg	tgaaatttgt	8160
gatgctattg	ctttatattg	aaccattata	agctgcaata	aacaagttgg	ggtggggcga	8220
gaactccagc	atgagatccc	cgcgctggag	gatcatccag	ccggcgtccc	ggaaaacgat	8280
tcggaagccc	aacctttcat	agaaggcggc	ggtggaatcg	aaatctcgta	gcacgtgtca	8340
gtctgtctcc	tcggccacga	agtgcacgca	gttgccggcc	gggtcgcgca	gggcgaactc	8400
ccgccccac	ggctgtctcg	cgatctcggg	catggccggc	ccggaggcgt	cccgggaagt	8460
cgtggacacg	acctccgacc	actcggcgta	cagctcgtcc	aggccgcgca	cccacaccca	8520
ggccagggtg	ttgtccggca	ccacctgggtc	ctggaccgcg	ctgatgaaca	gggtcacgtc	8580
gtcccggacc	acacgggca	agtcgtcctc	cacgaagtcc	cgggagaacc	cgagccggtc	8640
ggtccagaac	tcgaccgctc	cggcgacgtc	gcgcgcggtg	agcaccggaa	cggcactggg	8700
caacttggtc	atggatccag	atttcgctca	agttagtata	aaaaagcagg	cttcaatcct	8760
gcaggaattc	gatcgacact	ctcgtctact	ccaagaatat	caaagataca	gtctcagaag	8820
accaaagggc	tattgagact	tttcaacaaa	gggttaatat	gggaaacctc	ctcggattcc	8880
attgccagc	tatctgtcac	ttcatcaaaa	ggacagtaga	aaaggaagg	ggcacctaca	8940
aatgccatca	ttgcgataaa	ggaaaggcta	tcgttcaaga	tgctctgccc	gacagtgggc	9000
ccaaagatgg	acccccaccc	acgaggagca	tcgtggaaaa	agaagacgtt	ccaaccacgt	9060
cttcaaagca	agtggattga	tcgataaaca	tggtggagca	cgacactctc	gtctactcca	9120
agaatatcaa	agataacgtc	tcagaagacc	aaagggttat	tgagactttt	caacaaaggg	9180
taatatcggg	aaacctcctc	ggattccatt	gcccagctat	ctgtcacttc	atcaaaagga	9240
cagtagaaaa	ggaagggtgg	acctacaaat	gccatcattg	cgataaagga	aaggctatcg	9300
ttcaagatgc	ctctgccgac	agtgggtcca	aagatggacc	cccacccacg	aggagcatcg	9360
tggaaaaaga	agacgttcca	accacgtctt	caaagcaagt	ggattgatgt	gatattctca	9420
ctgacgtaa	ggatgacgca	caatcccact	atccttcgca	agaccttctt	ctatataagg	9480
aagttcattt	catttgagga	ggacacgctg	aaatcaccag	tctctctcta	caaatctatc	9540
tctctcgagc	tttcgcgag	ccgggggggg	aatgagatat	gaaaaagcct	gaactcaccg	9600
cgacgtctgt	cgagaagttt	ctgatcgaaa	agttcgacag	cgtctccgac	ctgatgcagc	9660
tctcggaggg	cgaagaatct	cgtgctttca	gcttcgatgt	aggaggcggt	ggatatgtcc	9720
tgccgggtaaa	tagctgcgcc	gatgggtttc	acaaagatcg	ttatgtttat	ggcacttttg	9780
catcgccgcg	gctcccgatt	ccggaagtgc	ttgacattgg	ggagtttagc	gagagcctga	9840
cctattgcat	ctcccgcgt	gcacagggtg	tcacgttgca	agacctgcct	gaaaccgaac	9900
tgcccgcgtg	tctacaaccc	gtcgcggagg	ctatggatgc	gatcgtgctg	gcccacttta	9960
gcccagcag	cgggttcggc	ccattcggag	cgcaaggaa	cgggtcaatac	actacatgga	10020
gtgatttcag	atgcgcgatt	gctgatcccc	atgtgtatca	ctggcaaaact	gtgatggacg	10080
acaccgtcag	tgcgtccgtc	gcgcaggctc	tcgatgagct	gatgcttttg	gcccaggact	10140
gcccgaagt	ccggcacctc	gtgcacgcgg	atttcggctc	caacaatgtc	ctgacggaca	10200
atggccgcag	aacacgggtc	attgactgga	gcgaggcgat	gttcggggat	tccaataacg	10260
aggtcgccaa	catcttcttc	tggaggcgtg	ggttggtttg	tatggagcag	cagacgcgtc	10320
acttcgagcg	gaggcatccg	gagcttgcag	gatcgccacg	actcggggcg	tatatgctcc	10380
gcattgggtc	tgaccaactc	tatcagagct	tggttgacgg	caatttcgat	gatgcagctt	10440
gggcgcaggg	tcgatgcgac	gcaatcgtcc	gatccggagc	cgggactgtc	gggcgtacac	10500

-70-

aaatcgcccc	cagaagcgcg	gccgtctgga	ccgatggctg	tgtagaagta	ctcgccgata	10560
gtggaaaccg	acgccccagc	actcgctccga	gggcaaagaa	atagagtaga	tgccgaccgg	10620
atctgtcgat	cgacaagctc	gagtttctcc	ataataatgt	gtgagtagtt	cccagataag	10680
ggaattaggg	ttcctatagg	gtttcgctca	tgtgttgagc	atataagaaa	cccttagtat	10740
gtatttgtat	ttgtaaaata	cttctatcaa	taaaatttct	aattcctaaa	acaaaaatcc	10800
agtactaaaa	tccagatccc	ccgaattaat	tcggcgttaa	ttcagatcaa	gcttgacctg	10860
gaatatcgcg	agtaaaactga	aaatcacgga	aaatgagaaa	tacacacttt	aggacgtgaa	10920
atatggcgag	gaaaactgaa	aaaggtggaa	aatttagaaa	tgtccactgt	aggacgtgga	10980
atatggcaag	aaaactgaaa	atcatggaaa	atgagaaaaca	tccacttgac	gacttgaaaa	11040
atgacgaaat	cactaaaaaa	cgtgaaaaat	gagaaatgca	cactgaagga	ctccgcggga	11100
attcgattgt	gctagccaat	gtttaacaag	atgtcaagca	caatgaatgt	tggtgggttg	11160
tggtcgtggc	tggtgggtgg	ggaaaattgc	ggtggttcga	gcggtagtga	tcggcgatgg	11220
ttggtgtttg	cagcgggtgt	tgatatcgga	atcacttatg	gtggttgtca	caatggaggt	11280
gcgtcatggg	tattggtggg	tggtcatcta	tatatTTTTA	taataatatt	aagtatttta	11340
cctatttttt	acatatTTTT	tattaaattt	atgcattggt	tgtatttttt	aatagttttt	11400
atcgtaactg	ttttataaaa	tattttatta	ttttatggtg	tatatattta	cttgatgtat	11460
tggaattttt	ctccattggt	ttttctatat	ttataataat	tttcttattt	ttttttgttt	11520
tattatgtat	tttttcgttt	tataataaat	atttattaaa	aaaaatatta	tttttgtaaa	11580
atataatcatt	tacaatgttt	aaaagtcatt	tgtgaatata	ttagctaagt	tgtacttctt	11640
tttgtgcatt	tggtgttgta	catgtctatt	atgattctct	ggccaaaaca	tgtctactcc	11700
tgctcacttg	gttttttttt	ttaaagacata	atcactagtg	attatatcta	gactgaaggc	11760
gggaaacgac	aatctgatca	tgagcggaga	attaagggag	tcacgttatg	accccgcccg	11820
atgacgcggg	acaagccggt	ttacgtttgg	aactgcacaga	accgcaacgt	tgaaggagcc	11880
actcagcccg	gggtttctgg	agtttaatga	gctaagcaca	tacgtcagaa	accattattg	11940
cgcggtcaaa	agtcgcctaa	ggctactatc	agctagcaaa	tatttcttgt	caaaaaatgc	12000
ccactgacgt	tccataaatt	ccctcggta	tccaattaga	gtctcatatt	cactctcaat	12060
ccaaataatc	tgcaccggat	ctcgagatcg	aattcccgcg	gcgcgcaatt	cactagtgga	12120
tccccgggta	cggtcagtc	cttatgttac	gtcctgtaga	aaccccaacc	cgtgaaatca	12180
aaaaactcga	cggcctgtgg	gcattcagtc	tggtatcgga	aaactgtgga	attgagcagc	12240
gttggtggga	aagcgcgtta	caagaaagcc	gggcaattgc	tgtgccaggc	agttttaacg	12300
atcagttcgc	cgatgcagat	attcgttaatt	atgtgggcaa	cgtctggtat	cagcgcgaag	12360
tctttatacc	gaaaggttgg	gcaggccagc	gtatcgtgct	gcgtttcgat	gcggtcactc	12420
attacggcaa	agtgtgggtc	aataatcagg	aagtgtgga	gcatacgggc	ggctatacgc	12480
catttgaagc	cgatgtcacg	ccgtatgta	ttgccgggaa	aagtgtacgt	atcacagttt	12540
gtgtgaacaa	cgaactgaac	tggcagacta	tcccgccggg	aatggtgatt	accgacgaaa	12600
acggcaagaa	aaagcagtc	tacttccatg	atcttcttaa	ctacgccggg	atccatcgca	12660
gcgtaatgct	ctacaccacg	ccgaacacct	gggtggacga	tatcaccgtg	gtgacgcgat	12720
tcgcgcaaga	ctgtaaccac	gcgtctgttg	actggcagg	gggtggccaat	gggtgatgtca	12780
gcgttgaact	gcgtgatgcg	gatcaacagg	tggttgcaac	tggacaaggc	accagcggga	12840
ctttgcaagt	ggtgaatccg	cacctctggc	aaccgggtga	aggttatctc	tatgaactgt	12900
acgtcacagc	caaaagccag	acagagtgtg	atatctaccc	gctgcgcgtc	ggcatccggg	12960
cagtgggcag	gaagggcgaa	cagttcctga	tcaaccacaa	accgttctac	tttactgggt	13020
ttggccgtca	tgaagatgcg	gatttgcgcg	gcaaaggatt	cgataacgtg	ctgatgggtg	13080
acgataccgc	attaatggac	tggattgggg	ccaactccta	ccgtacctcg	cattaccctt	13140
acgctgaaga	gatgctcgac	tgggcagatg	aacatggcat	cgtggtgatt	gatgaaactg	13200
cagctgtcgg	ctttaaccct	tctttaggca	ttggtttcga	agcgggcaac	aagccgaaag	13260
aactgtacag	cgaagaggca	gtcaacgggg	aaactcagca	ggcgacttta	caggcgatta	13320
aagagctgat	agcgcgtgac	aaaaaccacc	caagcgtggg	gatgtggagt	attgccaacg	13380
aaccggatac	ccgtccgcaa	ggtgcacggg	aatattttcg	gccactggcg	gaagcaacgc	13440
gtaaactcga	tccgacgcgt	ccgatcacct	gcgtcaatgt	aatgttctgc	gacgctcaca	13500
ccgataccat	cagcgatctc	tttgatgtgc	tgtgcctgaa	ccgttattac	ggttggtatg	13560
tccaaagcgg	cgatttggaa	acggcagaga	aggtactgga	aaaagaactt	ctggccctgg	13620
aggagaaact	gcatacagcc	attatcatca	ccgaatacgg	cgtggatacg	ttagccgggg	13680
tgcaactcaat	gtacaccgac	atgtggagtg	aagagtatca	gtgtgcatgg	ctggatatgt	13740
atcacccgct	ctttgatcgc	gtcagcgcgc	tcgtcgggtg	acaggtagtg	aatttcgcgc	13800
atttttgcgac	ctcgcaaggc	atattgcgcg	ttggcggtaa	caagaagggg	atcttcaccc	13860
gcgacgcgaa	accgaagtgc	gcggcttttc	tgctgcaaaa	acgctggact	ggcatgaact	13920
tcggtgaaaa	accgcgcag	ggaggcaaac	aatgaattca	caactctcct	ggcgaccat	13980
cgtcggttac	agcctcggga	attgcgtacc	gagctcgaat	ttccccgatc	gttcaaacat	14040
ttggcaataa	agttttctta	gattgaatcc	tgttgcgggt	cttgcgatga	ttatcatata	14100
atttctgttg	aattacgtta	agcatgtaat	aattaacatg	taatgcata	cgttatttat	14160
gagatggggt	tttatgatta	gagtcgccga	attatacatt	taatacgcga	tagaaaaaca	14220
aatatagcgc	gcaaaactag	ataaaattat	gcgcgcgggt	tcactctatg	tactagatcg	14280
ggaattcgat	atcaagcttg	gcactggccg	tcgttttaca	acgtcgtgac	tgggaaaacc	14340
ctggcggttac	ccaacttaat	cgccttgacg	cacatccccc	tttcgccagc	tggcgtaata	14400
gcgaagaggc	ccgcaccgat	cgccttcccc	aacagtttgc	cagcctgaat	ggcgaatgct	14460
agagcagctt	gagcttggat	cagattgtcg	tttcccgcc	tcagttttaa	ctatcagtg	14520

-71-

ttgacaggat atattggcgg gtaaacctaa gagaaaagag cgtttattag aataacggat 14580
 atttaaaagg gcgtgaaaag gtttatccgt tcgtccattt gtatgtg 14627

<210> 110
 <211> 9080
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> p18attBZeo(6XHS4)2eGFP Plasmid

<400> 110
 cagttgccgg ccgggtcggc cagggcgaac tccccccccc acggtctgct gccgatctcg 60
 gtcattggccg gcccgaggc gtcccggaag ttctgtggaca cgacctcga ccactcggcg 120
 tacagctcgt ccaggccggc caccacacc caggccaggg tgttgtcgg caccacctgg 180
 tcctggaccg cgtgatgaa cagggtcac tcgtcccgga ccacaccggc gaagtcgtcc 240
 tccacgaagt cccgggagaa cccgagccgg tcggtccaga actcgaccgc tccggcgacg 300
 tcgcgcggcg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
 caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatata gaattcctgc 420
 agccccgcgg atccgctcac ggggacagcc ccccccaaaa gccccaggg atgtaattac 480
 gtccctcccc cgttagggg cagcagcgag ccgcccgggg ctccgctccg gtccggcgct 540
 ccccccgcat atccccgag cggcagcgtg gggacagccc gggcacgggg aaggtggcac 600
 gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca cctgggggat 660
 acggggccgc ggatccgctc acggggacag cccccccca aagccccag ggatgtaatt 720
 acgtccctcc ccgctaggg ggcagcagcg agccgcccgg ggctccgctc cggtcggggc 780
 ctccccccgc atccccgag cggcagcgtg cggggacagc ccgggcacgg ggaaggtggc 840
 acgggatcgc ttctctctga acgcttctcg ctgctctttg agcctgcaga cacctggggg 900
 atacggggcc gcggatccgc tcacggggac agcccccccc caaagcccc agggatgtaa 960
 ttacgtccct cccccgctag ggggcagcag cgagccgccc ggggctccgc tccggctccg 1020
 cgctcccccc gcatccccga gccggcagcg tcgggggaca gccggggcac ggggaagggt 1080
 gcacgggata gctttctctc gaacgcttct cgctgctctt tgagcctgca gacacctggg 1140
 ggatacgggg ccgcgatcc gctcacgggg acagcccccc cccaaagccc ccagggatgt 1200
 aattacgtcc ctcccccgct agggggcagc agcagccgc ccggggctcc gctccggctc 1260
 ggcgctcccc ccgcatcccc gagccggcag cgtcggggga cagcccgggc acgggggaagg 1320
 tggcacggga tcgctttcct ctgaacgctc ctgctgctc tttgagcctg cagacacctg 1380
 ggggatacgg ggcgcgggat ccgctcacgg ggagacccc cccccaaagc cccagggat 1440
 gtaattacgt ccctcccccg ctagggggca gcagcagacc gcccggggt ccgctccggg 1500
 ccggcgctcc ccccgcatcc ccgagccggc agcgtgcggg gacagcccg ggacggggaa 1560
 ggtggcacgg gatcgcttcc ctctgaacgc ttctcgctgc tctttgagcc tgcagacacc 1620
 tgggggatac ggggcgcgg atccgctcac ggggacagcc ccccccaaaa gccccaggg 1680
 atgtaattac gtccctcccc cgctaggggg cagcagcgag ccgcccgggg ctccgctccg 1740
 gtccggcgct ccccccgcat ccccgagccg gcagcgtgc gggacagccc gggcacgggg 1800
 aaggtggcac gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca 1860
 cctgggggat acggggcggg ggatccacta gttattaata gtaatcaatt acgggggtc 1920
 tagttcatag cccatatatg gaggttccgc ttacataact tacggtaaat ggcccgctg 1980
 gctgaccgcc caacgacccc cggccattga cgtcaataat gacgtatgtt cccatagtaa 2040
 cgccaatagg gactttccat tgacgtcaat ggggtgacta tttacggtaa actgccact 2100
 tggcagtaca tcaagtgtat catatgccaa gtacgcccc tatttgacgtc aatgacggta 2160
 aatggcccg ctaggcattat gccagtaga tgaccttat ggactttcct acttggcagt 2220
 acatctacgt attagtcata gctattacca tgggtcgagg tgagccccc gttctgcttc 2280
 actctcccca tctccccccc ctccccaccc ccaattttgt atttatttat tttttaatta 2340
 ttttgctgag cgatgggggg gggggggggg ggggcgcgc ccaggcgggg cggggcgggg 2400
 cgagggggcg ggcggggcga ggcgggaggg tcgggcggca gccaatcaga gcggcgcgt 2460
 ccgaaaagttt ccttttatgg cgaggcgcg ggcggcgcg ccctataaaa agcgaagcg 2520
 gcggcgggcg ggagtcgctg cgttgccctc ccccgtgccc ccgctccgag ccgcctcgcg 2580
 ccgcccgcgc cggctctgac tgaccgcgtt actcccacag gtgagcgggc gggacggccc 2640
 ttctctcccg ggctgtaatt agcgtttggg ttaatgacgg ctcgcttctt ttctgtggct 2700
 gcgtgaaagc cttaaagggc tccgggaggg ccttttgctc gggggggagg ggctcggggg 2760
 gtgcgtgctg gtgtgtgtgc gtggggagcg ccgctgctgc ccgcgctgc 2820
 tgagcgctgc gggcgcggg cggggctttg tgcgctccgc gtgtgcgcga ggggagcgcg 2880
 gccggggcg gtgccccgc gtgcgggggg gctgcgagg gaacaaaagg ggtcgggctg 2940
 gtgtgtgctg ggggggggtg gcaggggggt tgggcgcggc ggtcgggctg taaccccc 3000
 ctgcaccccc ctccccaggt tctgagcac gggccggctt cgggtgcggg gctccgtgcg 3060
 gggcgtaggc cggggctcgc cgtgcgggc ggggggtggc ggcaggtagg ggtgcggg 3120
 ggggcggggc cgcctcgggc cggggagggc tcgggggagg ggcgcggcg ccccgagcg 3180
 ccggcggtcg tcgaggcgcg gcgagccgca gccattgcct tttatggtaa tcgtgcgaga 3240
 gggcgaggg acttcctttg tcccaaatct ggcgagccg aaatctggga ggcgcgcgcg 3300

-72-

cacccccctct	agcggggcgcg	ggcgaagcgg	tgcggcgccg	gcaggaagga	aatggggcggg	3360
gagggccttc	gtgcgtcgcc	gcgccgcgct	ccccttctcc	atctccagcc	tcggggctgc	3420
cgcaggggga	cggtgcctt	cggggggggac	ggggcagggc	gggggttcggc	ttctggcgctg	3480
tgaccggcgg	ctctagagcc	tctgctaacc	atgttcatgc	cttcttcttt	ttcctacagc	3540
tcctgggcaa	cgtgctgggt	gttgtgctgt	ctcatcattt	tggcaaagaa	ttcgccacca	3600
tgggtgagcaa	gggcgaggag	ctgttcacgc	gggtggtgcc	catcctgggt	gagctggagc	3660
gcgacgtaaa	cggccacaag	ttcagcgtgt	cggcgagggg	cgagggcgat	gccacctacg	3720
gcaagctgac	cctgaagttc	atctgcacca	ccggcaagct	gcccgtgccc	tggcccaccc	3780
tcgtgaccac	cctgacctac	ggcgtgcagt	gcttcagccg	ctaccccgac	cacatgaagc	3840
agcacgactt	cttcaagtcc	gccatgcccg	aaggctacgt	ccaggagcgc	accatcttct	3900
tcaaggacga	cggcaactac	aagaccgcgc	ccgaggtgaa	gttcgagggc	gacacctgtg	3960
tgaacgcgat	cgagctgaag	ggcatcgact	tcaggaggga	cggcaacatc	ctggggcaca	4020
agctggagta	caactacaac	agccacaacg	tctatatcat	ggccgacaag	cagaagaacg	4080
gcatcaaggt	gaacttcaag	atccgccaca	acatcgagga	cggcagcgtg	cagctcgccg	4140
accactacca	gcagaacacc	cccctcgccg	acggccccgt	gctgctgccc	gacaaccact	4200
acctgagcac	ccagtccgcc	ctgagcaaa	accccaacga	gaagcgcgat	cacatggctc	4260
tgctggagtt	cgtgaccgcc	gccgggatca	ctctcgccat	ggacgagctg	tacaagtaag	4320
aattcactcc	tcaggtgcag	gctgcctatc	agaaggtggg	ggctgggtgtg	gccaatgccc	4380
tggctcacaa	ataccactga	gatctttttc	cctctgccaa	aaattatggg	gacatcatga	4440
agcccccttga	gcattctgact	tctggcta	aaaggaaatt	tattttcatt	gcaatagtgt	4500
gttggaaattt	tttgtgtctc	tcactcggaa	ggacatatgg	gagggcaa	catttaaaac	4560
atcagaatga	gtatttgggt	tagagtttgg	caacatatgc	catatgctgg	ctgccatgaa	4620
caaagggtggc	tataaagagg	tcattcagtat	atgaaacagc	cccctgctgt	ccattcctta	4680
ttccatagaa	aagccttgac	ttgaggttag	atttttttta	tattttgttt	tgtgttattt	4740
ttttcttttaa	catccctaaa	attttcctta	catgttttac	tagccagatt	tttctctctc	4800
tcctgactac	tcaccgtcat	agctgtccct	cttctcttat	gaagatccct	cgacctcgag	4860
cccaagcttg	catcgactga	ggcgcactct	agtggatccc	ccgccccgta	tccccagggt	4920
gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	agcgatcccc	tgccaccttc	4980
cccggtgccc	ggctgtcccc	gcacgctgcc	ggctcgggga	tgcgggggga	gcgcgggacc	5040
ggagcgggagc	cccgggcggc	tcgctgctgc	cccctagcgg	gggagggagc	taattacatc	5100
cctggggggt	ttgggggggg	gctgtccccg	tgagcggatc	cgcggggccc	tatccccag	5160
gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	aaagcgatcc	cgtgccacct	5220
tccccgtgcc	cgggctgtcc	ccgcacgctg	ccggctcggg	gatgcggggg	gagcgccgga	5280
ccggagcggga	gccccggggc	gctcgtgtgt	gccccctagc	gggggagggga	cgtaattaca	5340
tccttggggg	ctttgggggg	gggctgtccc	cgtgagcggga	tcgcggggccc	cgatcccccc	5400
agggtgtctgc	aggctcaaa	agcagcgaga	agcgttccaga	ggaaagcgat	cccggtccac	5460
cttccccgtg	ccggggctgt	ccccgcacgc	tgccggctcg	gggatgcggg	gggagcgccg	5520
gaccggagcg	gagccccggg	cggctcgtgt	ctgcccccta	gcgggggagg	gacgtaatta	5580
catccctggg	ggctttgggg	gggggctgtc	cccgtagcgc	gatccgcggc	cccgtatccc	5640
ccaggtgtct	gcaggctcaa	agagcagcga	gaagcgttca	gaggaaagcg	atcccgtgcc	5700
accttccccg	gtccccgcac	gtccccgcac	gctcccggct	cggggatgcg	gggggagcgc	5760
cggaccggag	cggagccccg	ggcggtctgc	tgctgcccc	tagcggggga	gggacgta	5820
tacatccctg	ggggcttttg	gggggggctg	tcctcgtag	cggatccgcg	gccccgtatc	5880
ccccaggtgt	ctgcaggctc	aaagagcagc	gagaagcgtt	cagaggaaag	cgatcccgtg	5940
ccacctcccc	cgtgccccg	acgctcgcgc	acgctcgcgc	ctcggggatg	cggggggagc	6000
gccggaccgg	agcggagccc	cgggcggctc	tgctgctgcc	cctagcgggg	gagggagcga	6060
attacatccc	tgggggcttt	gggggggggc	tgtccccgtg	agcggatccg	cggccccgta	6120
tccccaggt	gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	agcgatcccc	6180
tgccaccttc	cccggtgccc	ggctgtcccc	gcacgctgcc	ggctcgggga	tgcgggggga	6240
gcgcgggaac	ggagcggagc	cccgggcggc	tcgctgctgc	cccctagcgg	gggagggagc	6300
taattacatc	cctgggggct	ttgggggggg	gctgtccccg	tgagcggatc	cgcggggctg	6360
caggaattcg	taatcatggt	catagctgtt	tcctgtgtga	aattgttatc	cgctcacaat	6420
tccacacaac	atacgagccg	gaagcataaa	gtgtaaagcc	tgggggtgct	aatgagttag	6480
ctaactcaca	ttaattgcgt	tgcgtcact	gcccgttttc	cagtcgggga	acctgtcgtg	6540
ccagctgcgt	taatgaatcg	gccaacgcgc	gggggagagg	ggtttgctga	ttggggcctc	6600
ttccgcttcc	tcgctcactg	actcgtgcg	ctcggtcgtt	cggctgcggc	gagcggatc	6660
agctcactca	aaggcggtaa	tacggttatc	cacagaatca	ggggataacg	caggaaagaa	6720
catgtgagca	aaaggccagc	aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggt	6780
tttccatagg	ctgacgagca	ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagaggtg	6840
gcgaaacccg	acaggactat	aaagatacca	ggcgtttccc	cctggaagct	ccctcgtgcg	6900
ctctcctgtt	ccgacctgc	cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	6960
cgtggcgctt	tctcatagct	cacgctgtag	gtatctcagt	tcgggtgtagg	tcgttcgctc	7020
caagctgggc	tgtgtgcacg	aacccccgt	tcagcccgac	cgtgctgctc	tatccggtaa	7080
ctatcgtctt	gagtcacaac	cggtaagaca	cgacttatcg	ccactggcag	cagccactgg	7140
taacaggatt	agcagagcga	ggatgttagg	cgggtgtaca	gagttcttga	agtgggtggc	7200
taactacggc	tacactagaa	ggacagtatt	tggtatctgc	gctctgctga	agccagttac	7260
cttcggaaaa	agagttggta	gctcttgatc	cggcaaaaaa	accaccgctg	gtagcgggtg	7320

-73-

```

tttttttgggt  tgcaagcagc  agattacgcg  cagaaaaaaa  ggatctcaag  aagatccttt  7380
gatcttttct  acgggggtctg  acgctcagtg  gaacgaaaaa  tcacgttaag  ggatttttgg  7440
catgagatta  tcaaaaagga  tcttcaccta  gatcctttta  aattaaaaat  gaagttttta  7500
atcaatctaa  agtatatatg  agtaaaacttg  gtctgacagt  taccaatgct  taatcagtga  7560
ggcacctatc  tcagcgatct  gtctatctcg  ttcatccata  gttgcctgac  tccccgtcgt  7620
gtagataact  acgatacggg  agggccttacc  atctggcccc  agtgctgcaa  tgataccgcg  7680
agacccacgc  tcaccggctc  cagattttatc  agcaataaac  cagccagccg  gaagggccga  7740
gcgagagaag  gggtcctgcaa  ctttatccgc  ctccatccag  tctattaatt  gttgcccggg  7800
agctagagta  agtagttcgc  cagttaatag  tttgcgcaac  gttgttgcca  ttgctacagg  7860
catcggtggg  tcacgctcgt  cgtttggtat  ggcttcattc  agctccgggt  cccaacgatc  7920
aaggcgagtt  acatgatccc  ccatggtgtg  caaaaaagcg  gttagctcct  tcggctcctc  7980
gatcgttgtc  agaagtaagt  tggccgcagt  gttatcactc  atggttatgg  cagcactgca  8040
taattctctt  actgtcatgc  catccgtaag  atgcttttct  gtgactgggt  agtactcaac  8100
caagtcattc  tgagaatagt  gtatgcggcg  accgagttgc  tcttgcccg  cgtcaatacg  8160
ggataatacc  gcgccacata  gcagaacttt  aaaagtgtc  atcattggaa  aacgttcttc  8220
ggggcgaaaa  ctctcaagga  tcttaccgct  gttgagatcc  agttcgatgt  aaccactcgt  8280
tgcacccaac  tgatcttcag  catcttttac  tttcaccagc  gtttctgggt  gagcaaaaac  8340
aggaaggcaa  aatgccgcaa  aaaagggaat  aagggcgaca  cggaaatgtt  gaatactcat  8400
actcttcctt  tttcaatatt  attgaagcat  ttatcagggt  tattgtctca  tgagcggata  8460
catatttgaa  tgtattttag  aaaataaaca  aatagggtt  ccgcgcacat  ttccccgaaa  8520
agtgccacct  gacgtagtta  acaaaaaaaa  gccgccgaa  gcgggcttta  ttaccaggcg  8580
aagcgccatt  cgccattcag  gctgcgcaac  tgttgggaag  ggcgatcggt  gcgggcctct  8640
tcgctattac  gccagctggc  gaaaggggga  tgtgctgcaa  ggcgattaag  ttgggtaacg  8700
ccagggtttt  ccagtcacg  acgttgtaaa  acgacggcca  gtccgtaata  cgactcactt  8760
aaggccttga  ctagagggtc  gacgggtatac  agacatgata  agatacattg  atgagtttgg  8820
acaaaccaca  ctagaatgc  agtgaaaaaa  atgctttatt  tgtgaaattt  gtgatgctat  8880
tgcttttttt  gtaaccatta  taagctgcaa  taaacaagtt  ggggtgggcg  aagaactcca  8940
gcatgagatc  cccgcgctgg  aggatcatcc  agccggcgtc  ccggaaaacg  attccgaagc  9000
ccaacctttc  atagaaggcg  gcggtggaat  cgaaatctcg  tagcacgtgt  cagtcctgct  9060
cctcggccac  gaagtgcacg  9080

```

<210> 111

<211> 4223

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attBBSRpolyA10 Plasmid

<400> 111

```

gttaactacg  tcagggtggca  cttttcgggg  aaatgtgcgc  ggaaccccta  tttgtttatt  60
tttctaaata  cattcaaata  tgtatccgct  catgagacaa  taacctgat  aaatgcttca  120
ataatattga  aaaaggaaga  gtatgagtat  tcaacatttc  cgtgtcgccc  ttattccctt  180
ttttgcggca  ttttgccttc  ctgttttttg  tcaccagaa  acgctgggtg  aagtaaaaga  240
tgctgaagat  cagttgggtg  ctcgagtggt  ttacatcgaa  ctggatctca  acagcggtaa  300
gatccttgag  agtttttcgc  ccgaagaacg  ttctccaatg  atgagcactt  ttaaagtctt  360
gctatgtggc  gcggtattat  cccgtgttga  cgccgggcaa  gagcaactcg  gtccgcgcgt  420
acactattct  cagaatgact  tgggtgagta  ctccacagtc  acagaaaagc  atcttacgga  480
tggcatgaca  gtaagagaat  tatgcagtgc  tgccataacc  atgagtata  acactgcggc  540
caacttactt  ctgacaacga  tcggaggacc  gaaggagcta  accgcttttt  tgcacaacat  600
gggggatcat  gtaactcgcc  ttgatcggtt  ggaaccggag  ctgaatgaag  ccataccaaa  660
cgacgagcgt  gacaccacga  tgcctgtagc  aatggcaaca  acgttgcgca  aactattaac  720
tggcgaaact  ctactctag  cttcccgcca  acaattaata  gactggatgg  aggccgataa  780
agttgcagga  ccacttctgc  gctcggccct  tccggctggc  tggtttattg  ctgataaaat  840
tggagccggt  gagcgtgggt  ctgcgggtat  cattgcagca  ctggggccag  atggtaagcc  900
ctcccgatc  gtagttatct  acacgacggg  gagtcaggca  actatggatg  aacgaaatag  960
acagatcgct  gagataggtg  cctcactgat  taagcattgg  taactgtcag  accaagttta  1020
ctcatatata  cttagatttg  atttaccctg  gttgataatc  agaaaagccc  caaaaacagg  1080
aagattgtat  aagcaaatat  ttaaattgta  aacgttaata  ttttggttaa  attcgcgtta  1140
aatttttgtt  aaatcagctc  attttttaac  caataggccg  aaatcggcaa  aatcccttat  1200
aatcaaaag  aatagcccg  gatagggttg  agtggtgttc  cagtttgcaa  caagagttca  1260
ctattaaaga  acgtggactc  caacgtcaaa  gggcgaaaaa  ccgtctatca  gggcgatggc  1320
ccactacgtg  aaccatcacc  caaatcaagt  tttttggggt  cgaggtgccc  taagcacta  1380
aatcggaacc  ctaaggagag  cccccgattt  agagcttgac  ggggaaagcg  aacgtggcga  1440
gaaaggaagg  gaagaaagcg  aaaggagcgg  gcgctagggc  gctggcaagt  gtacgggtca  1500
cgctgcgctg  aaccaccaca  cccgccgcgc  ttaatgcgcc  gctacagggc  gcgtaaaagg  1560
atctaggtga  agatcctttt  tgataatctc  atgacaaaaa  tcccttaacg  tgagttttcg  1620

```

-74-

```

ttccactgag cgtcagaccc cgtagaaaaa atcaaaggat cttcttgaga tccttttttt 1680
ctgocgcgtaa tctgctgctt gcaaacaaaa aaaccaccgc taccagcggg ggtttgtttg 1740
ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata 1800
ccaaatactg ttctttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca 1860
ccgcctacat acctcgctct gctaatacctg ttaccagtgg ctgctgccag tggcgataag 1920
tcgtgtctta ccgggttggg ctcaagacga tagttaccgg ataaggcgca gcggctgggc 1980
tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 2040
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aaggagagaa ggcggacagg 2100
tatccggtaa gcggcagggg cggaacagga gagcgacga gggagcttcc aggggggaaac 2160
gcctgggtatc tttatagtcc tgtcgggttt cgccacctct gacttgagcg tcgatttttg 2220
tgatgctcgt cagggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg 2280
ttcctggcct tttgctggcc ttttgctcac agttaatgtg agttagctca ctctattaggc 2340
accccaggct ttacacttta tgcttccggc tcgtatgttg tgtggaattg tgagcggata 2400
acaatttcac acaggaaaca gctatgacca tgattacgcc aagctacgta atacgactca 2460
ctagtggggc ccgtgcaatt gaagccggct ggcgccaagc ttctctgcag gattgaagcc 2520
tgctttttta tactaacttg agcgaaatct ggtaccatc gaaaacattt aacatttctc 2580
aacaagatct agaattagta gaagtagcga cagagaagat tacaatgctt tatgaggata 2640
ataaacatca tgtgggagcg gcaattcgtg cgaaaacagg agaatcatt tcggcagtag 2700
atattgaagc gtatatagga cgagtaactg tttgtgcaga agccattgag attggtagtg 2760
cagtttcgaa tggacaaaag gattttgaca cgattgtagc tgttagacac ccttattctg 2820
acgaagtaga tagaagtatt cgagtggtaa gctcttggg tatgtgtagg gagtgtattt 2880
cagactatgc accagattgt tttgtgttaa tagaaatgaa tggcaagtta gtcaaaacta 2940
cgattgaaga actcattcca ctcaaataa cccgaaatta aaagttttac cataccaagc 3000
ttggctgctg cctgaggctg gacgacctcg cggagtctta ccggcagtg ccaatccgtcg 3060
gcatccagga aaccagcagc ggctatccgc gcacccatgc ccccgactg caggagtggg 3120
gaggcacgat ggcgctttg gtcggatct ttgtgaagga accttacttc tgtggtgtga 3180
cataattgga caaactacct acagagattt aaagctctaa ggtaaataa aaatttttaa 3240
gtgtataatg tgttaacta ctgattctaa ttgtttgtg atttttagatt ccaacctatg 3300
gaactgatga atgggagcag tgggtggaatg cctttaatga ggaaaacctg ttttgctcag 3360
aagaaatgcc atctagtgt gatgaggcta ctgctgactc tcaacattct actcctccaa 3420
aaaagaagag aaaggtagaa gaccccaagg accttccttc agaattgcta agttttttga 3480
gtcatgctgt gtttagtaat agaactcttg cttgctttgc tatttacacc acaaaggaaa 3540
aagctgcaat gctatacaag aaaattatgg aaaaatattc tgtaaccttt ataagtaggc 3600
ataacagtta taatcataac atactgtttt ttcttactcc acacaggcat agagtgtctg 3660
ctattaataa ctatgtcaa aaattgtgta cctttagctt tttaatttgt aaaggggtta 3720
ataaggaata tttgatgtat agtgcttga ctagagatca taatcagcca taccactttt 3780
gtagaggttt tacttgcttt aaaaaacctc ccacacctcc cctgaacctt gaaacataaa 3840
atgaatgcaa ttgttggtgt taacttgttt attgcagctt ataatgggtta caaataaagc 3900
aatagcatca caaatttcac aaataaagat ccacgaattc gctagcttcg gccgtgacgc 3960
gtctccggat gtacaggcat gcgtcgacc ctagtcaag gcttaagtg agtcgtatta 4020
cggactggcc gtctgtttac aacgtcgtga ctgggaaaaa cctggcggtta cccaacttaa 4080
tcgccttgca gcacatcccc ctttcgccag ctggcgtaat agcgaagagg cccgcaccga 4140
tcgccttcc caacagttgc gcagcctgaa cgtctcgctt cgcttcgctt ggtaataaag 4200
cccgcttcgg cgggcttttt ttt 4223

```

<210> 112

<211> 5855

<212> DNA

<213> Artificial Sequence

<220>

<223> pCX-LamIntR Plasmid

<400> 112

```

gtcgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60
gcccatatat ggagttccgc gttacataac ttacggtaaa tggccgcgct ggctgaccgc 120
ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180
ggactttcca ttgacgtcaa tgggtggact atttacggta aactgcccac ttggcagtag 240
atcaagtgtg tcatatgcca agtacgcccc ctattgacgt caatgacggg aaatggccc 300
cctggcatta tgcccagtag atgaccttat gggactttcc tacttggcag tacatctacg 360
tattagtcat cgctattacc atgggtcgag gtgagcccca cgttctgctt cactctcccc 420
atctcccccc cctccccacc cccaattttg tattttattt ttttttaatt attttgtgca 480
gcgatggggg cggggggggg gggggcgcg gccaggcggg gcggggcggg gcgagggggc 540
gggcggggcg aggcggagag gtgcggcgcc cttattgacgt agcggcgcgc tccgaaagt 600
tccttttatg gcgagggcgg ggcggcgggc gccctataaa aagcgaagcg cgcggcgggc 660
gggagtcgct gcgttgccct cgcgccgtgc cccgctccgc gccgcctcgc gccggccgcc 720
cgggctctga ctgaccgcgt tactcccaca ggtgagcggg cgggacggcc cttctcctcc 780

```

-75-

gggctgtaat	tagcgcttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgcgtgaaag	840
ccttaaagg	ctccgggagg	gccctttgtg	cgggggggag	cggctcgggg	gggtcggtgcg	900
tgtgtgtgtg	cgtgggggagc	gccgcgtgcg	gcccgcgtcg	cccggcgggt	gtgagcgctg	960
cgggcgcggc	gcggggcctt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggccggggggc	1020
ggtgccccgc	ggtgcggggg	ggctgcgagg	ggaacaaagg	ctgcgtgcgg	ggtgtgtgcg	1080
tgggggggtg	agcagggggg	gtgggcgcgg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cggcccggct	tcgggtgcgg	ggctccgtgc	ggggcgctggc	1200
gcggggctcg	ccgtgcgggg	cgggggggtg	cggcaggtgg	gggtgccggg	cggggcgggg	1260
ccgcctcggg	ccggggaggg	ctcgggggag	gggcgcggcg	gccccggagc	gcccggcggt	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttatggta	atcgtgcgag	agggcgcgagg	1380
gacttccttt	gtcccaaata	tggcggagcc	gaaatctggg	aggcgccgcc	gcaccccc	1440
tagcgggcgc	gggcgaagcg	gtgcggcgcc	ggcaggaagg	aaatgggccc	ggagggcctt	1500
cgtgcgtcgc	cgcgcgcgcg	tccccttctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acggctgcct	tcggggggga	cggggcaggg	cgggggttcg	cttctggcgt	gtgaccggcg	1620
gctctagagc	ctctgctaac	catgttcatg	ccttcttctt	tttctacag	ctcctgggca	1680
acgtgctggg	tgttgtgctg	tctcatcatt	ttggcaaaaga	attcatggga	agaaggcgaa	1740
gtcatgagcg	ccgggattta	ccccctaacc	tttatataag	aaacaatgga	tattactgct	1800
acagggaccc	aaggacgggt	aaagagtttg	gattaggcag	agacaggcga	atcgcaatca	1860
ctgaagctat	acaggccaac	attgagttat	tttcaggaca	caaacacaag	cctctgacag	1920
cgagaatcaa	cagtataat	tccgttacgt	tacattcatg	gcttgatcgc	tacgaaaaaa	1980
tcttggccag	cagaggaatc	aagcagaaga	cactcataaa	ttacatgagc	aaaattaaag	2040
caataaggag	gggtctgcct	gatgctccac	ttgaagacat	caccacaaaa	gaaattgcgg	2100
caatgctcaa	tggatacata	gacgagggca	aggcggcgct	agccaagtta	atcagatcaa	2160
cactgagcga	tgcattccga	gaggcaatag	ctgaaggcca	tataacaaca	aaccatgtcg	2220
ctgccactcg	cgcagcaaaa	tctagagtaa	ggagatcaag	acttacggct	gacgaatacc	2280
tgaataattt	tcaagcagca	gaatcatcac	catgtttggc	cagacttgca	atggaaatcg	2340
ctgttgttac	cgggcaacga	gttggtgatt	tatgcgaaat	gaagtggctc	gatatcgtag	2400
atggatatct	ttatgtcgag	caaagcaaaa	caggcgtaaa	aattgccatc	ccaacagcat	2460
tgcataattga	tgtctcggga	atatcaatga	aggaaacact	tgataaatgc	aaagagattc	2520
ttggcggaga	aaccataatt	gcactactc	gtcgcgaacc	gctttcatcc	ggcacagtat	2580
caaggtatct	tatgcgcgca	cgaaaagcat	caggtctttc	cttcgaaggg	gagtcgccta	2640
cctttcacga	gttgcgagc	ttgtctgcaa	gactctatga	gaagcagata	agcgataagt	2700
ttgctcaaca	tcttctcggg	cataagtcgg	acaccatggc	atcacagtat	cgtgatgaca	2760
gaggcaggga	gtgggacaaa	attgaaatca	aataagaatt	cactcctcag	gtgcaggctg	2820
cctatcagaa	ggtggtggct	ggtgtggcca	atgccctggc	tcacaaaatac	cactgagatc	2880
tttttccctc	tgccaaaaat	tatggggaca	tcatagaagc	ccttgagcat	ctgacttctg	2940
gctaataaag	gaaatttatt	ttcattgcaa	tagtgtgttg	gaattttttg	tgtctctcac	3000
tcggaaggac	atatgggagg	gcaaatcatt	taaaacatca	gaatgagtat	ttggtttaga	3060
gtttggaac	atatgccata	tgtctgctgc	catgaacaaa	ggtggctata	aagaggtcat	3120
cagtagatga	aatagccctc	tgctgtccat	tctttattcc	atagaaaagc	cttgacttga	3180
ggtagatttt	tttttatatt	ttgttttgtg	ttattttttt	ctttaacatc	cctaaaaattt	3240
tccttacatg	ttttactagc	cagatttttc	ctcctctcct	gactactccc	agtcatagct	3300
gtccctcttc	tcttatgaag	atccctcgac	ctgcagccca	agcttggcgt	aatcatggct	3360
atagctgttt	cctgtgtgaa	attgttatcc	gtccacaatt	ccacacaaca	tacgagccgt	3420
aagcataaag	tgtaaagcct	gggggtgccta	atgagtgagc	taactcacat	taattgcgtt	3480
gcgtcactg	ccgcctttcc	agtcgggaaa	cctgtcgtgc	cagcggatcc	gcatctcaat	3540
tagtcagcaa	ccatagtcct	gcccctaact	cgcgccatcc	cgcccctaac	tccgcccagt	3600
tccgcccatt	ctccgcccc	tggttgacta	atttttttta	tttatgcaga	ggccgaggcc	3660
gcctcggcct	ctgagctatt	ccagaagtag	tgaggaggct	tttttggagg	cctaggcttt	3720
tgcaaaaaagc	taacttgttt	ataatggtta	tgcatcttag	caaataaagc	aatagcatca	3780
caaatttcac	aaataaagca	ttttttttcac	tgcatcttag	ttgtgggttg	tccaaactca	3840
tcaatgtatc	ttatcatgtc	tggatccgct	gcattaatga	atcggccaac	gcgcggggag	3900
aggcggtttg	cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgcgtcgggt	3960
cgttcggctg	cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	4020
atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaagg	cagcaaaaagg	ccaggaaccg	4080
taaaaaggcc	gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacaa	4140
aaatcgacgc	tcaagtcaga	ggtggcgaaa	cccgacagga	ctataaagat	accaggcggt	4200
tccccctgga	agctccctcg	tgcgctctcc	tggtccgacc	ctgccgctta	ccggatacct	4260
gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcaa	tgctcacgct	gtaggtatct	4320
cagttcggtg	taggtcgctc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	4380
cgaccgctgc	gccttatccg	gtaactatcg	tcttgagtcg	aaccgggtaa	gacacgactt	4440
atcgccactg	gcagcagcca	ctggtaacag	gatttagcaga	gcgagggtatg	taggcgggtgc	4500
tacagagttc	ttgaagtggg	ggcctaacta	cggctacact	agaaggacag	tatttgggtat	4560
ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	4620
acaaaccacc	gctggttagcg	gtggtttttt	tgtttgcaag	cagcagatta	cgcgcagaaa	4680
aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaaacya	4740
aaactcacgt	taagggattt	tggctcatgag	attatcaaaa	aggatcttca	cctagatcct	4800

-76-

```

tttaaattaa aaatgaagtt ttaaataaat ctaaagtata tatgagtaaa cttgggtctga 4860
cagttaccac tgcttaatac gtgaggcacc tatctcagcg atctgtctat ttcgttcac 4920
catagttgcc tgactccccg tcgtgtagat aactacgata cgggagggct taccatctgg 4980
ccccagtgct gcaatgatac cgcgagaccc acgctcaccg gctccagatt tatcagcaat 5040
aaaccagcca gccggaaggc cggagcgagc aagtggtcct gcaactttat ccgcctccat 5100
ccagttctatt aattgttgcc gggaagctag agtaagtagt tcgccagtta atagtttgcc 5160
caacgttggtt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc 5220
attcagctcc ggttcccaac gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa 5280
agcgggttagc tccttcggtc ctccgatcgt tgtcagaagt aagttggccg cagtgttatc 5340
actcatgggt atggcagcac tgcataattc tottactgtc atgccatccg taagatgctt 5400
ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc ggcgaccgag 5460
ttgtctttgc ccggcgtaaa tacgggataa taccgcgcca catagcagaa ctttaaaagt 5520
gctcatcatt ggaaaacgtt ctccggggcg aaaactctca aggatcttac cgctgttgag 5580
atccagttcg atgtaaccac ctctgtgcacc caactgatct tcagcatctt ttactttcac 5640
cagcgtttct gggtagcaaa aaacaggaag gcaaaaatgcc gcaaaaaagg gaataagggc 5700
gacacggaaa tgttgaatac tcatactctt cctttttcaa tattattgaa gcatttatca 5760
gggttattgt ctcatgagcg gatacatatt tgaattgtatt tagaaaaata aacaaatagg 5820
ggttcgcgcg acatttcccc gaaaagtgcc acctg 5855

```

<210> 113

<211> 4346

<212> DNA

<213> Artificial Sequence

<220>

<223> pSV40-193AttpsensePur Plasmid

<400> 113

```

ccggtgccc caccatcccc tgaccacgc cctgacccc tcacaaggag acgaccttcc 60
atgaccgagt acaagccac ggtgcgcctc gccaccgcg acgacgtccc cggggccgta 120
cgcacccctc cgcgcgcgtt cgcgactac ccgccacgc gccacaccgt cgaccggac 180
cgccacatcg agcgggtcac cgagctgcaa gaactcttcc tcacgcgcgt cgggctcgac 240
atcggaagg tgtgggtcgc ggacgacggc gccgcggtgg cggctctggac cacgccggag 300
agcgtcgaag cggggggcgtt gttcgcgag atcgccccgc gcatggccga gttgagcgg 360
tcccggttgg ccgcgcagca acagatggaa ggcctcctgg cgccgcaccg gcccaaggag 420
cccgctgggt tcctggccac cgtcggcgct gcccccacc accagggcaa gggctctggg 480
agcgcgctcg tgctccccgg agtgaggcg gccgagcgc ccgggggtgc cgccttctcg 540
gagacctcgc cgccccgcaa cctccccctc tacgagcggc tcggcttcac cgtcacccgc 600
gacgtcgagg tgcccgaagg accgcgcacc tggtgcatga cccgcaagcc cggtgctga 660
cgcccccccc acgacccgca cgcgccgacc gaaaggagcg caccacccca cggctccgac 720
cgaagccgac cggggcgccc ccgcgcaccc ccgagggccc accgactcta 780
gaggatcata atcagccata ccacatttgt agaggtttta cttgctttaa aaaacctccc 840
acacctcccc ctgaacctga aacataaaat gaatgcaatt gttgttgtaa acttgtttat 900
tgcagcttat aatggttaca aataaagcaa tagcatcaca aatttcacaa ataaagcatt 960
tttttctatg cattctagtt gtggtttgtc caaactcatc aatgtatctt atcatgtctg 1020
gatccgcgac ggatccttaa ttaagtctag agtcgactgt ttaaacctgc aggcattgca 1080
gcttggcgta atcatggtca tagctgtttc ctgtgtgaaa ttgttatccg ctcaaatc 1140
cacacaacat acgagccgga agcataaagt gtaaaagcct ggggtgcctaa tgagttagct 1200
aactcacatt aattgcgttg cgtcactgc ccgctttcca gtcgggaaac ctgctggtcc 1260
agctgcatta atgaatcggc caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt 1320
ccgcttcttc gctcactgac tcgtgcgct cggctcgttc gctgcggcga gcggtatcag 1380
ctcactcaaa ggcggttaata cggttatcca cagaatcagg ggataacgca ggaaagaaca 1440
tgtgagcaaa aggcagcaa aaggccagga accgtaaaaa ggccgcgttg ctggcgtttt 1500
tccataggct ccgccccctt gacgagcatc acaaaaatcg acgctcaagt cagaggtggc 1560
gaaacccgac aggaactata agataccagg cgtttcccc tggaagctcc ctctgctgct 1620
ctcctgttcc gacctgccg cttaccggat acctgtccgc ctttctccct tcgggaagcg 1680
tggcgcttcc tcatagctca cgctgtaggt atctcagttc ggtgtaggtc gttcgtcca 1740
agctgggctg tgtgcacgaa cccccgttc agcccgaccg ctgocctta tccggtaact 1800
atcgtcttga gtccaaccc gtaagacacg acttatcgcc actggcagca gccactggta 1860
acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag tgggtggccta 1920
actacggcta cactagaagg acagtatttg gtatctgcgc tctgctgaag ccagttacct 1980
tcggaaaaag agttggtagc tcttgatccg gcaaaacaaac caccgctggg agcgggtgg 2040
tttttgtttt caagcagcag attacgcgca gaaaaaagg atctcaagaa gatcctttga 2100
tctttcttac ggggtctgac gtcagtga acgaaaactc acgttaaggg attttgggtc 2160
tgagattatc aaaaaggatc ttcacctaga tcctttttaa ttaaaaatga agtttttaaa 2220
caatctaaag tatatatgag taaacttggt ctgacagtta ccaatgctta atcagtgagg 2280
cacctatctc agcgatctgt ctatttcgtt catccatagt tgccgtactc ccgctcgtgt 2340

```


-77-

```

agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag 2400
accacgcgtc accggctcca gatttatcag caataaacca gccagccgga agggccgagc 2460
gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag 2520
ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tggtgccatt gctacaggca 2580
tcgttggtgtc acgctcgtcg tttggtatgg cttcattcag ctccggttcc caacgatcaa 2640
ggcgagttac atgatcccc atgttggtgca aaaaagcggg tagctccttc ggtcctccga 2700
tcgttggtcag aagtaagttg gccgcagtgt taccactcat gggtatggca gcactgcata 2760
attctcttac tgtcatgccca tccgtaagat gcttttctgt gactggtgag tactcaacca 2820
agtcattctg agaatagttg atgcggcgac cgagttgctc ttgcccggcg tcaatacggg 2880
ataataccgc gccacatagc agaactttaa aagtgtcat cattggaaaa cgttcttcgg 2940
ggcgaaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa ccactcgtg 3000
caccacaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaacag 3060
gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatac 3120
tcttcctttt tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggataca 3180
tatttgaaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt ccccgaaaag 3240
tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa aataggcgta 3300
tcacgagggc ctttcgtctc gcgcgtttcg gtgatgacgg tgaaaacctc tgacacatgc 3360
agctcccgga gacggtcaca gcttgtctgt aagcggatgc cgggagcaga caagcccgtc 3420
agggcgcgtc agcgggtgtt ggcggtgtgc ggggctgggt taactatgcg gcatcagagc 3480
agattgtact gagagtgcac catatgcggt gtgaaatacc gcacagatgc gtaaggagaa 3540
aataccgcac caggcgccat tcgccattca ggctgcgcaa ctggtgggaa gggcgatcgg 3600
tgccggcctc ttcgctatta cgccagctgg cgaaaagggg atgtgctgca aggcgattaa 3660
gttgggtaac gccagggttt tcccagtcac gacgttgtaa aacgacggcc agtgaattcg 3720
agctgtggaa tgtgtgtcag ttagggtgtg gaaagtcccc aggtccccca gcaggcagaa 3780
gtatgcaaaag catgcatctc aattagtcag caaccagggtg tggaaagtcc ccaggctccc 3840
cagcaggcag aagtatgcaa agcatgcac tcaattagtc agcaaccata gtcccgcccc 3900
taactccgac catcccgccc ctaactccgc ccagttccgc ccattctccg ccccatggct 3960
gactaatttt ttttatttat gcagaggccg aggcgcgctc ggctctctgag ctattccaga 4020
agtagtgagg aggccttttt ggaggctcgg taccoccttg cgctaattgt ctgttacagg 4080
tcactaatac catctaagta gttgattcat agtgactgca tatgttgtgt tttacagtat 4140
tatgtagtct gttttttatg caaaatctaa tttaatatat tgatatttat atcatattac 4200
gtttctcggt cagctttttt atactaagtt ggcattataa aaaagcattg cttatcaatt 4260
tggtgcaacg aacaggtcac tatcagtc aaataaatca ttatttgatt tcaattttgt 4320
cccactccct gcctctgggg ggcgcg 4346

```

<210> 114

<211> 3166

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo Plasmid

<400> 114

```

cagttgccgg ccgggtcgcg cagggcgaac tccgcgcccc acgggtgctc gccgatctcg 60
gtcatggcgg gcccgaggc gtcccgaag ttctgtggaca cgacctcga ccactcggcg 120
tacagctcgt ccaggccgcg caccacacc caggccaggg tggtgtccgg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacacgggc gaagtctgcc 240
tccacgaagt cccgggagaa cccgagccgg tcggtccaga actcgaccgc tccggcgagc 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaagca ggcttcaatc ctgcagagaa gcttgcatgc ctgcaggctg 420
actctagagg atccccgggt accgagctcg aattcgtaat catggtcata gctgtttcct 480
gtgtgaaatt gttatccgct cacaattcca cacaacatac gagccggaag cataaagtgt 540
aaagcctggg gtgcctaatt agtgagctaa ctcacattaa ttgcgttgcg ctactgccc 600
gctttccagt cgggaaacct gtcgtgccag ctgcattaat gaatcggcca acgcgcgggg 660
agaggcgggt tgcgtatttg gcgctcttcc gcttccctgc tcaactgact gctgcgctcg 720
gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtaataac gttatccaca 780
gaatcagggg ataacgcagg aaagaacatg tgagcaaaa ggcagcaaaa ggccaggaa 840
cgtaaaaagg ccgcgttgct ggcgtttttc ataggctcc gccccctga cgagcatcac 900
aaaaatcgac gctcaagtca gaggtggcga aaccgcagag gactataaag ataccaggcg 960
tttccccctg gaagctccct cgtgcgctct cctgttccga cctgcccgt ataccggatac 1020
ctgtccgcct ttctcccttc ggggaagcgt gcgctttctc atagctcacg ctgtagggtat 1080
ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag 1140
ccgacccgct gcgccttate cggtaactat cgtcttgagt ccaaccgggt aagacacgac 1200
ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta tgtaggcggg 1260
gtacagaggt tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttggt 1320
atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc 1380

```

-78-

```

aaacaaacca ccgctggttag cgggtggtttt tttgtttgca agcagcagat tacgcgcaga 1440
aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc tcagtggaac 1500
gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt cacctagatc 1560
cttttaaat aaatatgaag ttttaaatca atctaaagta tatatgagta aacttggtct 1620
gacagttacc aatgcttaat cagttaggca cctatctcag cgatctgtct atttcgttca 1680
tccatagtgt cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct 1740
ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca 1800
ataaaccagc cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc 1860
atccagtcta ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtgtg 1920
cgcaacggtt ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggatgggt 1980
tcattcagct ccggttccca acgatcaagg ctagttacat gatcccccat gttgtgcaaa 2040
aaagcgggta gtccttcgg tcctccgatc gtgtgcagaa gtaagttggc cgcagtgtta 2100
tactcatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaatgtgc 2160
ttttctgtga ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg 2220
agttgctctt gcccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa 2280
gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg 2340
agatccagtt cgtgtaacc cactcgtgca ccaactgat cttcagcatc ttttactttc 2400
accagcgttt ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg 2460
gcgacacgga aatggtgaat actcatactc ttcttttttc aatattattg aagcatttat 2520
cagggttatt gtctcatgag cggatacata ttggaatgta tttagaaaaa taaacaaata 2580
gggggtccgc gcacatttcc ccgaaaagtg ccactgacg tagttaacaa aaaaaagccc 2640
gccgaagcgg gctttattac caagcgaagc gccattcgcc attcaggctg cgcaactgtt 2700
gggaagggcg atcgggtcgg gcctcttcgc tattacgcca gctggcgaaa gggggatgtg 2760
ctgcaaggcg attaatgttg gtaacgccag ggttttccca gtcacgacgt tgtaaaacga 2820
cggccagtc gtaatacgac tcaactaagg ccttgactag agggtcgacg gtatacacag 2880
atgataagat acattgatga gtttgacaa accacaacta gaatgcagtg aaaaaaatgc 2940
tttatttgtg aaatttgtga tgctattgct ttatttgtaa ccattataag ctgcaataaa 3000
caagttgggg tgggcgaaga actccagcat gagatccccg cgctggagga tcatccagcc 3060
ggcgtcccg aaaacgattc cgaagcccaa ctttcatag aaggcggcgg tggaatcgaa 3120
atctcgtagc acgtgtcagt cctgctcttc gccacgaag tgcacg 3166

```

<210> 115

<211> 7600

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo3' 6XHS4eGFP Plasmid

<400> 115

```

cagttgccgg ccgggtcggc cagggcgaac tccgcggccc acggctgctc gccgatctcg 60
gtcatggcgg gcccgaggc gtcccggaag ttcgtggaca cgacctcga ccactcggcg 120
tacagctcgt ccaggccggc caccacacc caggccaggg tggtgtccgg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtgcctc 240
tccacgaagt cccgggagaa cccgagccgg tccgtccaga actcgaccgc tccggcgacg 300
tcgcgcggcg tgagcacggc aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaagca ggcttcaatc ctgcagagaa gcttgatcta gttattaata 420
gtaatcaatt acggggtcat tagttcatag ccatatatg gagttccgcg ttacataact 480
tacggtaaat gggccgctg gctgaccgcc caacgacccc cgccattga cgtcaataat 540
gacgtatggt cccatagtaa cgccaatagg gactttccat tgacgtcaat ggggtggacta 600
tttacggtaa actgcccact tggcagtaca tcaagtgtat catatgccaa gtacgcccc 660
tattgacgtc aatgacggta aatggcccgc ctggcattat goccagtaca tgaccttatg 720
ggactttcct acttggcagt acatctacgt attagtcac tctccccccc ctccccacc ccaattttgt 840
tgagccccac gttctgcttc acttccccca cgtctctgac cgtctctgac agcgttgggt ttaatgacgg 1200
atttatttat tttttaatta ttttgtgcag cgatgggggc gggggggggg ggggcgcgcg 900
ccaggcgggg cggggcgggg cgagggcgcg ggccggggcg ggcggagagg tgcggcgggc 960
gccaatcaga ggggcgcgct ccgaaagttt ggagtcgctg cgttgccctc gccccgtgcc 1020
ccctataaaa agcgaagcgc gcggcgggcg cggctctgac cgtctctgac tgaccgcgtt actcccacg 1140
cgcctcgcg cgcctcgcg ttctcctccg ggctgtaatt tccgggaggg ccctttgtgc 1260
gtgagcgggc gggacggccc gctgaaaagc gtgctgctgc gtgtgtgtgc gtggggagcg ccgctgctgc 1320
ctcgtttcct ttctgtggct gctgaaagc ggtgctgac gggcgcgcg ggggctcttg tgcgtccgc 1380
gggggggagc ggctcggggg tgagcgtcgc gggggggggc gggggggggg gctgcgaggg 1440
cccgcgctgc cgggcggctg ggggcggcgc gggggggggg gggggggggg ggggcgcggc 1500
gtgtgcgcga ggggagcgcg gccggggggc gggggggggg gggggggggg gggcgcggtt 1560
gaacaaaggc tgcgtgcggg gtgtgtgctg ctcccagat tgcgtgagca gggccggtt 1620
ggtcgggctg taaccccccc ctgcaccccc cggggtcgcg cgtgccccgc ggggggtggc

```

ggcaggtggg	ggtgccgggg	ggggcggggg	cgccctgggg	cggggagggg	tcgggggagg	1680
ggcgcgggcg	ccccggagcg	ccggcgggctg	tcgagggcgcg	gcgagccgca	gccattgcct	1740
tttatggtaa	tcgtgcgaga	gggcgcaggg	acttcctttg	tcccaaactct	ggcggagccg	1800
aaatctggga	ggcgccgccc	cacccctctct	agcgggcgcg	ggcgaagcgg	tgcggcgccc	1860
gcaggaagga	aatgggcggg	gagggcccttc	gtgcgtcgcc	gcgcgcgcgt	cccttctctc	1920
atctccagcc	tcgggggctgc	cgcaggggga	cggctgcctt	cgggggggac	ggggcagggc	1980
ggggtttcgg	ttctggcgctg	tgaccggcgg	ctctagagcc	tctgctaacc	atgttcatgc	2040
cttcttcttt	ttcctacagc	tcctgggcaa	cgtgctgggt	gttgtgctgt	ctcatcattt	2100
tggcaaagaa	ttcgccacca	tggtagcaaa	gggcgaggag	ctgttcaccg	gggtggtgcc	2160
catcctgggtc	gagctggacg	gcgacgtaaa	cggccacaag	ttcagcgtgt	ccggcgaggg	2220
cgaggcgcat	gccacctacg	gcaagctgac	ctgaagtctc	atctgcacca	ccggcaagct	2280
gccccgtccc	tggtccaccc	tcgtgaccac	cctgacctac	ggcgtgcagt	gcttcagccg	2340
ctaccccgac	cacatgaagc	agcacgactt	cttcaagtcc	gccatgcccg	aaggctacgt	2400
ccaggagcgc	accatcttct	tcaaggacga	cggcaactac	aagacccgcg	ccgaggtgaa	2460
gttcgagggc	gacacctctg	tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	2520
cggcaacatc	ctgggggcaca	agctggagta	caactacaac	agccacaacg	tcctatatcat	2580
ggccgacaag	cagaagaacg	gcatcaaggt	gaacttcaag	atccgccaca	acatcgagga	2640
cggcagcgtg	cagctcgccg	accactacca	gcagaacacc	cccatcgccg	acggccccgt	2700
gctgctgccc	gacaaccact	acctgagcac	ccagtccgcc	ctgagcaaa	accccaacga	2760
gaagcgcgat	cacatggtcc	tgctggagtt	cgtgaccgcc	gccgggatca	ctctcggcat	2820
ggacgagctg	ctgccatgaa	aattcaactcc	tcaggtgcag	gctgcctatc	agaagggtgt	2880
ggctggtgtg	gccaatgccc	tggtccacaa	ataccactga	gatctttttc	cctctgccaa	2940
aaattatggg	gacatcatga	agcccttga	gcctctgact	tctggctaata	aaaggaaatt	3000
tattttcatt	gcaatagtgt	gttggaattt	tttgtgtctc	tcactcgaa	ggacatatgg	3060
gagggcaaat	cattttaaac	atcagaatga	gtatttgggt	tagagtgttg	caacatatgc	3120
catatgctgg	ctggcatgaa	caaagggtgg	tataaagagg	tcatacgtat	atgaaacagc	3180
ccctgctgtg	ccattcctta	ttccatagaa	aagccttgac	ttgaggttag	atttttttta	3240
tattttgttt	tgtgttattt	ttttctttta	catccctaaa	attttcctta	catgttttac	3300
tagccagatt	tttcctcctc	tcctgactac	tccagctcat	agctgtccct	cttctcttat	3360
gaagatccct	cgacctgcag	cccaagcttg	cctgctgca	ggtcgactct	agtggatccc	3420
ccgccccgta	tcctccaggt	gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	3480
agcgatcccc	tgccaccttc	cccgtgccc	ggctgtcccc	gcacgctgcc	ggctcgggga	3540
tgcgggggga	gcgcgcggac	ggagcggagc	cccgggcggc	tcgctgctgc	cccctagcgg	3600
gggaggggag	taattacatc	cctgggggct	ttgggggggg	gctgtccccg	tgagcggatc	3660
cgcgcccccg	tatccccacg	gtgtctgcag	gtcctaaagag	cagcgagaag	cgttcaagag	3720
aaagcgatcc	cgtgccacct	tccccgtgcc	cgggctgtcc	ccgcacgctg	ccggctcggg	3780
gatgcggggg	gagcgccgga	ccggagcggg	gccccggggc	gctcgctgct	gccccctagc	3840
gggggagggg	cgtaattaca	tccttggggg	ctttgggggg	gggctgtccc	cgtgagcggg	3900
tcgcgcggccc	cgtatcccc	aggtgtctgc	aggctcaaag	agcagcgaga	agcgttcaga	3960
ggaaagcgat	ccctgcccac	cttccccgtg	cccggtgtgt	ccccgcacgc	tgccggctcg	4020
gggatgcggg	gggagcgccc	gaccggagcg	gagccccggg	cggctcgctg	ctgcccccta	4080
gcggggggag	gacgtaatta	catccctggg	ggctttgggg	gggggctgtc	cccgtgagcg	4140
gatccgcggc	cccgatatccc	ccaggtgtct	gcaggctcaa	agagcagcga	gaagcgttca	4200
gaggaaagcg	atccccgtgc	accttccccg	tgccccgggt	gtccccgcac	gctgctgccc	4260
cggggaatcg	ggggagcgcg	cggacgggag	cggaaccccg	ggcggtctcg	tgctgcccc	4320
tagcggggga	gggacgtaat	tacatccctg	ggggctttgg	gggggggctg	tccccgtgag	4380
cggatccgcg	gccccgtatc	ccccaggtgt	ctgcaggctc	aaagagcagc	gagaagcgtt	4440
cagaggaaag	cgatcccgtg	ccaccttccc	cgtgccccgg	ctgtccccgc	acgctgccgg	4500
ctcggggatg	cggggggagc	gccggaccgg	agcggagccc	cgggcggctc	gctgctgccc	4560
cctagcgggg	gagggacgta	attacatccc	tgggggcttt	gggggggggc	tgtccccgtg	4620
agcggatccg	cggccccgta	tccccaggt	gtctgcaggc	tcaaagagca	gcgagaagcg	4680
ttcagaggaa	agcgatcccc	tgccaccttc	cccgtgcccc	ggctgtcccc	gcacgctgcc	4740
ggctcggggg	tgcgggggga	gcgcgcggac	ggagcggagc	cccgggcggc	tcgctgctgc	4800
cccctagcgg	gggagggacg	taattacatc	cctgggggct	ttgggggggg	gctgtcccc	4860
tgagcggatc	cgcggggctg	caggaattcg	taatcatggg	catagctgtt	tcctgtgtga	4920
aattgtttatc	cgctcacaa	tccacacaac	atacgagccg	gaagcataaa	gtgtaaagcc	4980
tggggtgcct	aatgagttag	ctaactcaca	ttaattgcgt	tgcgctcact	gcccgttttc	5040
cagtcgggaa	acctgtcgtg	ccagctgcac	taatgaatcg	gccaacgcgc	ggggagaggg	5100
ggtttgcgta	ttcgcttccc	ttcgcttccc	tcgctcactg	actcgctcgc	ctcggtcgtt	5160
cggctgcggc	gagcgggtatc	agctcactca	aaggcggtaa	tacggttatc	cacagaatca	5220
ggggataacg	caggaaagaa	catgtgagca	aaaggccagc	aaaaggccag	gaaccgtaaa	5280
aaggccgcgt	tgctggcggt	tttccatagg	ctccgcccc	ctgacgagca	tcacaaaaat	5340
cgacgctcaa	gtcagaggtg	gcgaaacccc	acagactat	aaagatacca	ggcgtttccc	5400
cctggaagct	ccctcgtgcg	ctctcctggt	ccgacctgc	cgcttaccgg	atacctgtcc	5460
gcctttctcc	cttcgggaag	cgtggcgctt	tctcatagct	cacgctgtag	gtatctcagt	5520
tcggtgtagg	tcgttcgctc	caagctgggc	tgtgtgcacg	aacccccgt	tcagccccgac	5580
cgctgcgcct	tatccggtaa	ctatcgtctt	gagtcacaac	cggtaaagaca	cgacttatcg	5640

-80-

ccactggcag	cagccactgg	taacaggatt	agcagagcga	ggatgtagg	cggtgctaca	5700
gagttcctga	agtgggtggc	taactacggc	tacactagaa	ggacagtatt	tggatctctgc	5760
gctctgctga	agccagttac	cttcggaaaa	agagttggta	gctcttgatc	cggcaaacaa	5820
accaccgctg	gtagcgggtg	tttttttggt	tgcaagcagc	agattacgcg	cagaaaaaaa	5880
ggatctcaag	aagatccttt	gatcttttct	acggggctctg	acgctcagtg	gaacgaaaaac	5940
tcacgttaag	ggattttggt	catgagatta	tcaaaaagga	tcttcaccta	gatccttttta	6000
aattaaaaat	ggtgcttttaa	atcaatctaa	agtatatatg	agtaaaacttg	gtctgacagt	6060
taccaatgct	taatcagtga	ggcacctatc	tcagcgatct	gtctatttcg	ttcatccata	6120
gttgccctgac	ccccgcctgt	gtagataact	acgatacggg	agggccttacc	atctggcccc	6180
agtgcctgcaa	tgataccgcg	agacccacgc	tcaccggctc	cagattttatc	agcaataaac	6240
cagccagccg	gaagggccga	gcgcagaagt	ggctcctgcaa	ctttatccgc	ctccatccag	6300
tctattaatt	gttgccggga	agctagagta	agtgttcgc	cagtttaatag	tttgcgcaac	6360
gttggttgcca	ttgctacagg	catcggtgtg	tcacgctcgt	cgtttggtat	ggcttcattc	6420
agctccggtt	cccaacgatc	aaggcgagtt	acatgatccc	ccatgttgtg	caaaaaagcg	6480
gttagctcct	tcggctcctc	gatcgttgtc	agaagtaagt	tggccgcagt	gttatcactc	6540
atgggttatgg	cagcactgca	taattctctt	actgtcatgc	catccgtaag	atgcttttct	6600
gtgactgggtg	agtactcaac	caagtcattc	tgagaatagt	gtatgcggcg	accgagttgc	6660
tcttgcccgg	cgtcaatacg	ggataataacc	gcgccacata	gcagaacttt	aaaagtgtct	6720
atcattggaa	aacgtttctc	ggggcgaaaa	ctctcaagga	tcttaccgct	gttgagatcc	6780
agttcgatgt	aaccactcgc	tgaccccaac	tgatcttcag	catcttttac	tttcaccagc	6840
gtttctgggt	gagcaaaaaac	aggaaggcaa	aatggccgcaa	aaaagggaat	aaggcgcaac	6900
cggaaatggt	gaatactcat	actcttcctt	tttcaatatt	attgaagcat	ttatcagggg	6960
tattgtctca	tgagcggata	catatttgaa	tgtatttaga	aaaataaaca	aataggggtt	7020
ccgcgcacat	ttccccgaaa	agtgcacact	gacgtagtta	acaaaaaaa	gcccccgcaa	7080
gcgggcttta	ttaccaagcg	aagcgccatt	cgccattcag	gctgcgcaac	tggtgggaag	7140
ggcgatcggt	tcgggcctct	tcgctattac	gccagctggc	gaaaggggga	tggtgctgcaa	7200
ggcgattaag	ttgggtaacg	ccagggtttt	cccagtcacg	acgttgtaaa	acgacggcca	7260
gtccgtaata	cgactcactt	aaggccttga	ctagagggtc	gacggtatac	agacatgata	7320
agatacattg	atgagtttgg	acaaaccaca	actagaatgc	agtgaaaaaa	atgctttatt	7380
tgtgaaattt	gtgatgctat	tgctttatatt	gtaaccatta	taagctgcaa	taaacaaagt	7440
ggggtggggc	aagaactcca	gcatgagatc	cccgcgctgg	aggatcatcc	agccggcgct	7500
ccggaaaacg	attccgaagc	ccaacctttc	atagaaggcg	gcggtggaat	cgaaatctcg	7560
tagcacgtgt	cagtcctgct	cctcggccac	gaagtgcacg			7600

<210> 116

<211> 7631

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo5'6XHS4eGFP Plasmid

<400> 116

cagttgcg	ccgggtcg	cagggcgaac	tccgcggc	acggctgctc	gccgatctcg	60
gtcatggcg	gcccggaggc	gtcccgaag	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggccg	cacccacacc	caggccaggg	tggtgtccgg	caccacctgg	180
tcctggaccg	cgctgatgaa	cagggtcacg	togtcccga	ccacaccggc	gaagtctctc	240
tcacgaagt	ccggggagaa	cccagccgg	tcggtccaga	actcgaccgc	tccggcgacg	300
tcgcgcgcgg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgatata	gaattctctc	420
agccccgcgg	atccgctcac	ggggacagcc	cccccccaa	gccccaggg	atgtaattac	480
gtccctcccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	gtccggcgct	540
ccccccgcat	ccccgagccg	gcagcgtg	gggacagccc	gggcacgggg	aagggtggcac	600
gggatcgctt	tcctctgaac	gctctctgct	gctctttgag	cctgcagaca	cctgggggat	660
acggggccgc	ggatccgctc	acggggacag	cccccccca	aagccccag	ggatgtaatt	720
acgtccctcc	cccgctaggg	ggcagcagcg	agccgcccgg	ggctccgctc	cggtccggcg	780
ctccccccgc	atccccgagc	cggcagcgtg	cggggacagc	ccgggacagg	ggaaggtggc	840
acgggatcgc	tttctctga	acgcttctcg	ctgctctttg	agcctgcaga	cacctggggg	900
atacggggcc	tcagggtccg	agccccccc	agccccccc	caaagcccc	agggatgtaa	960
ttacgtccct	cccccgctag	ggggcagcag	cgagccgccc	ggggctccgc	tccgggtccg	1020
cgctcccccc	gcacccccga	gccggcagcg	tgccggggaca	gccccgggac	ggggaaggtg	1080
gcacgggatc	gctttctctt	gaacgcttct	cgctgctctt	tgagcctgca	gacacctggg	1140
ggatacgggg	ccgcggatcc	gctcacgggg	acgcgcccc	cccaaagccc	ccagggtgat	1200
aattacgtcc	ctcccccgct	agggggcagc	agcgagccgc	ccggggctcc	gctccggtcc	1260
ggcgctcccc	ccgcaccccc	gagccggcag	cgtgcggggg	cagccccggg	acggggaagg	1320
tggcacggga	tcgctttcct	ctgaacgctt	ctcgctgctc	tttgagcctg	cagacacctg	1380
ggggatacgg	ggccgcggat	ccgctcacgg	ggacagcccc	ccccaaagc	ccccagggat	1440

-81-

gtaattacgt	ccctcccccg	ctaggggggca	gcagcgagcc	gcccgggggct	ccgctccgggt	1500
ccggcgctcc	ccccgcaccc	ccgagccggc	agcgtgcggg	gacagcccgg	gcacgggggaa	1560
gggtggcacg	gatcgctttc	ctctgaacgc	ttctcgctgc	tctttgagcc	tgcagacacc	1620
tgggggatac	ggggcccgcg	atccgctcac	ggggacagcc	cccccccaa	gcccccagg	1680
atgtaattac	gtccctcccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	1740
gtccggcgct	ccccccgcac	ccccgagccg	gcagcgctgc	gggacagccc	gggcacgggg	1800
aagggtggc	gggacgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttattaata	gtaatcaatt	acggggtcat	1920
tagttcatag	cccatatatg	gagttccgcg	ttacataact	tacggtaaat	ggcccgccctg	1980
gctgaccgcc	caacgacccc	cgcccataga	cgtaataaat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgcccact	2100
tggcagtaac	tcaagtgat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagtaac	tgaccttatg	ggacttttct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tgggtcgagg	tgagccccac	gttctgcttc	2280
actctcccca	tctccccccc	ctccccaccc	ccaattttgt	atatttttat	tttttaatta	2340
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cgggcgggg	2400
cgagggggcg	ggcgggggcg	ggcgagagg	tgccggcgga	gccaatcaga	gcggcgcgct	2460
ccgaaagt	ccttttatgg	cgaggcggg	gcggcgcgcg	ccctataaaa	agcgaagcgc	2520
gcgggggcg	ggagtcgctg	cgttgccttc	gccccgtg	ccgctccg	ccgctccg	2580
ccgcccgc	cggtctgac	tgaccgcgtt	actcccacag	gtgagcggg	gggacggccc	2640
ttctcctccg	ggctgtaatt	agcgccttgg	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
gcgtgcaag	cttaaagg	tccgggagg	ccctttgtgc	gggggggag	ggctcgggg	2760
gtgcgtgcgt	gtgtgtgtgc	gtggggagcg	ccgcgtgcgg	ccgcgctgc	ccggcgcgctg	2820
tgagcgctgc	gggcgcggcg	cggggctttg	tgcgctccgc	gtgtgcgcga	ggggagcgcg	2880
gcccggggcg	gtgccccgcg	gtgcgggggg	gctgcgaggg	gaacaaagg	tgcggtcggg	2940
gtgtgtgcgt	ggggggggga	gcagggggtg	tggcgcggg	ggtcgggctg	taaccccc	3000
ctgcaccccc	ctccccgagt	tgctgagcac	ggccccgctt	cggggtcggg	gctccgtgcg	3060
gggcgtggcg	cggggctcgc	cgtgccgggc	ggggggtgg	ggcaggtggg	ggtgccgggc	3120
ggggcggggc	cgccctcggg	cggggagggc	tcgggggagg	ggcgcgcgcg	ccccggagcg	3180
ccggcgggctg	tcgaggcgcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
gggcgcaggg	acttcctttg	tcccaaactc	ggcgagccg	aaatctggga	ggcgcccg	3300
cacccccctc	agcgggcgcg	ggcgaagcgg	tgccgcgcg	gcaggaagga	aatggcggg	3360
gagggccttc	gtgcgtgcgc	gcgcgcgcgt	ccctctctcc	atctccagcc	tcggggctgc	3420
cgcaggggga	cggtgccttc	cgggggggac	ggggcaggg	gggggtcggc	ttctggcgctg	3480
tgaccggcg	ctctagagcc	tctgctaacc	atgttcatgc	cttcttcttt	ttctcagagc	3540
tcctgggcaa	cgtgctgggt	gttgtgctgt	ctcatcattt	tggcaaagaa	ttcgccacca	3600
tgggtgagcaa	gggcgaggag	ctgttcaccg	gggtggtgcc	catcctgggt	gagctggacg	3660
gcgacgtaaa	cgccacacaag	ttcagcgtgt	ccggcgagg	cgagggcgat	gccacctacg	3720
gcaagctgac	cctgaagt	atctgcacca	ccggcaagct	gcccgtgccc	tggcccacc	3780
tcgtgaccc	cctgacctac	ggcgtgcagt	ccttcagccg	ctaccccgac	cacatgaagc	3840
agcacgactt	cttcaagtcc	gccaatgccc	aaggctacgt	ccaggagcgc	accatcttct	3900
tcaaggacga	cggaactac	aagaccgcg	ccgaggtgaa	gttcgagggc	gacaccctgg	3960
tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	cggaacatc	ctggggcaca	4020
agctggagta	caactacaac	agccacaacg	cttatatcat	ggccgacaag	cagaagaacg	4080
gcatcaaggt	gaacttcaag	atccgccaca	acatcgagga	cggcagcggt	cagctcgccg	4140
accactacca	gcagaacacc	cccatcggcg	acggccccgt	gctgctgccc	gacaaccact	4200
acctgagcac	ccagtcgcgc	ctgagcaaa	accccaacga	gaagcgcgat	cacatgggtc	4260
tgctggagtt	cgtgaccgcc	gcccggatca	ctctcggcac	ggacgagctg	tacaagtaag	4320
aattcactcc	tcaggtgcag	gctgcctatc	agaaggtggt	ggctgggtgtg	gccaatgcc	4380
tggctcacaa	ataccactga	gatctttttc	cctctgccaa	aaattatggg	gacatcatga	4440
agccccctga	gcatctgact	tctggctaatt	aaaggaaatt	tatttttcatt	gcaatagtgt	4500
gttggaaattt	tttgtgtctc	tactcggaa	ggacatatgg	gagggcaaat	cattttaaacc	4560
atcagaatga	gtatttgggt	tagagtttgg	caacatatgc	catatgctgg	ctgccatgaa	4620
caaagggtgg	tataaagagg	tcacagat	atgaaacagc	ccctgctgt	ccattcctta	4680
ttccatagaa	aagccttgac	ttgaggttag	atttttttta	tattttgttt	tgtgttattt	4740
ttttctttta	catccctaaa	attttcttta	catgttttac	tagccagatt	tttctctctc	4800
tcctgactac	tcccagtcac	agctgtccct	cttctcttat	gaagatccct	cgacctgcag	4860
cccaagcttg	catgcctgca	ggtcgactct	agaggatccc	cggttaccga	gctcgaattc	4920
gtaatcatgg	tcctagtgtg	ttctgtgtgt	aaattgttat	ccgctcacaa	ttccacacaa	4980
catacgagcc	ggaagcataa	agtgtaaagc	ctgggggtgcc	taatgagtga	gctaactcac	5040
attaattg	ttgcgctcac	tgcccgcctt	ccagtcggga	aacctgtcgt	gccagctgca	5100
ttaatgaatc	ggccaacgcg	cggggagagg	cggtttgcgt	attgggcgct	cttcgccttc	5160
ctcgctcact	gactcgctgc	gctcggtcgt	tcggctcgcg	cgagcggtat	cagctcactc	5220
aaaggcggt	atacggttat	ccacagaatc	aggggataac	gcaggaaaga	acatgtgagc	5280
aaaaggccag	caaaaggcca	ggaaccgtaa	aaaggccgcg	ttgctggcgt	ttttccatag	5340
gctccgcccc	cctgacgagc	atcacaaaaa	tcgacgctca	agtcagaggt	ggcgaaaccc	5400
gacaggacta	taaagatacc	aggcgtttcc	ccctggaagc	tcctcgtgc	gctctcctgt	5460

-82-

tccgaccctg	ccgcttaccg	gatacctgtc	cgccctttctc	ccttcgggaa	gcgtggcgct	5520
ttctcatagc	tcacgctgta	ggatatctcag	ttcgggtgtag	gtcgttcgct	ccaagctggg	5580
ctgtgtgcac	gaaccccccg	ttcagcccgga	ccgctgcgcc	ttatccggta	actatcgtct	5640
tgagtccaac	ccggtaagac	acgacttatc	gccactggca	gcagccactg	gtaacaggat	5700
tagcagagcg	aggtatgtag	gcggtgctac	agagttcttg	aagtgggtggc	ctaactacgg	5760
ctacactaga	aggacagtat	ttggatatctg	cgctctgctg	aagccagtta	ccttcggaaa	5820
aagagttggg	agctcttgat	ccggcaaaaca	aaccacccgt	ggtagcgggtg	gtttttttgt	5880
ttgcaagcag	cagattacgc	gcagaaaaaaa	aggatctcaa	gaagatcctt	tgatcttttc	5940
tacggggctc	gacgctcagt	ggaacgaaaa	ctcacgttaa	gggatttttg	tcatgagatt	6000
atcaaaaagg	atcttcacct	agatcctttt	aaattaaaaa	tgaagtttta	aatcaatcta	6060
aagtatatat	gagtaaactt	ggtctgacag	ttaccaatgc	ttaatcagtg	aggcacctat	6120
ctcagcgatc	tgtctatttc	gttcattccat	agttgectga	ctccccgtcg	tgtagataac	6180
tacgatacgg	gagggccttac	catctggccc	cagtgtcgca	atgataccgc	gagaccacag	6240
ctcacccggc	ccagattttat	cagcaataaaa	ccagccagcc	ggaagggccg	agcgcagaag	6300
tggctcctgca	actttatccg	cctccatcca	gtctattaat	tggtgcgggg	aagctagagt	6360
aagttagttcg	ccagtttaata	gtttgcgcaa	cgttgcttgc	attgctacag	gcacgtgggt	6420
gtcacgctcg	tcgttttggt	tggcttcatt	agctccgggt	ttcccaacgat	caaggcgagt	6480
tacatgatcc	cccatgttgt	gcaaaaaagc	ggttagctcc	ttcggtcctc	cgatcgttgt	6540
cagaagtaag	ttggcccgag	tgttatcact	catggttatg	gcagcactgc	ataattctct	6600
tactgtcatg	ccatccgtaa	gatgcttttc	tgtgactggg	gagtactcaa	ccaagtcatt	6660
ctgagaatag	tgtatgcggc	gaccgagttg	ctcttgcccg	gcgtcaatac	gggataatac	6720
cgcgccacat	agcagaactt	taaaagtgtc	catcattgga	aaacgttctt	cggggcgaaa	6780
actctcaagg	atcttaccgc	tggtgagatc	cagttcgatg	taaccactc	gtgcacccaa	6840
ctgatcttca	gcatctttta	ctttcaccag	cgtttctggg	tgagcaaaaa	caggaaggca	6900
aaatgccgca	aaaaagggaa	taaggcgac	acggaaatgt	tgaatactca	tactcttctc	6960
ttttcaatat	tattgaagca	tttatcaggg	ttattgtctc	atgagcggat	acataattga	7020
atgtatttag	aaaaataaac	aaataggggt	tccgcgcaca	tttccccgaa	aagtgccacc	7080
tgacgtagtt	aacaaaaaaa	agcccgccga	agcgggcttt	attaccaagc	gaagcgccat	7140
tgcgccattca	ggctgcgcaa	ctgttgggaa	gggcgatcgg	tgccggcctc	ttcgctatta	7200
cgccagctgg	cgaagggggg	atgtgctgca	aggcgattaa	gttgggtaac	gccaggggtt	7260
tcccagtcac	gacgttgtaa	aacgacggcc	agtcggtaat	acgactact	taaggccttg	7320
actagagggg	cgacgggtata	cagacatgat	aagatacatt	gatgagtttg	gacaaaccac	7380
aactagaatg	cagtgaaaaa	aatgctttat	ttgtgaaatt	tgtgatgcta	ttgctttatt	7440
tgtaaccatt	ataagctgca	ataaacaagt	tgggggtggc	gaagaactcc	agcatgagat	7500
cccgcgctg	gaggatcatc	cagccggcgt	ccgggaaaac	gattccgaag	cccaaccctt	7560
catagaagcg	ggcgggtggaa	tcgaaatctc	gtagcacgtg	tcagtcctgc	tcctcggcca	7620
cgaagtgcac	g					7631

<210> 117

<211> 4615

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo6XHS4 Plasmid

<400> 117

cagttgccgg	ccgggtcgcg	cagggcgaa	tcccgccccc	acggctgctc	gccgatctcg	60
gtcatggccg	gcccggaggc	gtcccggaag	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggcccg	caccacacac	caggccaggg	tggtgtccgg	caccacctgg	180
tcctggaccg	cgctgatgaa	cagggtcacg	tcgtcccggg	ccacaccggc	gaagtcgtcc	240
tccacgaagt	cccgggagaa	cccagaccgg	tcgggtccaga	actcgaccgc	tccggcgacg	300
tcgcgcgcgg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagtttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgcatgc	ctgcaggctc	420
actctagtgg	atcccccgcc	ccgtatcccc	cagggtgtctg	caggctcaaa	gagcagcgag	480
aagcgttcag	aggaaaagca	tcccgtgcca	ccttccccgt	gcccgggctg	tccccgcacg	540
ctgcccggctc	ggggatgcgg	ggggagcgcc	ggaccggagc	ggagccccgg	gcccgtcgct	600
gctgccccct	agcggggggg	ggacgtaatt	acatccctgg	gggctttggg	ggggggctgt	660
ccccgtgagc	ggatcccgcg	ccccgtatcc	cccagggtgc	tgagggtca	aagagcagcg	720
agaagcgttc	agaggaaaagc	gatcccgtgc	caccttcccc	gtgcccgggc	tgtccccgca	780
cgctgcccgg	tcgggggatgc	ggggggagcg	ccggaccggg	gcggagcccc	gggcggctcg	840
ctgctgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	ggggggggct	900
gtccccgtga	gcggatccgc	ggcccgtat	ccccagggtg	tctgcaggct	caaagacag	960
cgagaagcgt	tcagaggaaa	gcgatcccg	gccaccttcc	ccgtgcccgg	gctgtccccg	1020
cacgctgccc	gctcggggat	gcggggggg	cgccggaccg	gagcggagcc	ccgggcccgt	1080
cgctgctgcc	ccctagcggg	ggagggacgt	aattacatcc	ctgggggctt	tggggggggg	1140
ctgtccccct	gagcgggatcc	gcggccccgt	atcccccagg	tgtctgcagg	ctcaaagagc	1200

-83-

```

agcgagaagc gttcagagga aagcgatccc gtgccacctt ccccggtgcc gggctgtccc 1260
cgacagctgc cggctcgggg atgcgggggg agcgccggac cggagcggag ccccgggcgg 1320
ctcgctgctg ccccttagcg ggggaggggac gtaattacat ccctgggggc tttggggggg 1380
ggctgtcccc gtgagcggat ccgcgggcccc gtatccccca ggtgtctgca ggctcaaaga 1440
gcagcgagaa gcgttcagag gaaagcgatc ccggtgccacc ttccccgtgc cggggtgtgc 1500
cccgacagct gccgggtcgg ggatgcgggg ggagcgccgg accggagcgg agccccgggc 1560
ggctcgctgc tccccctag cgggggaggg acgtaattac atccctgggg gctttggggg 1620
ggggtgtgtc ccgtgagcgg atccgcgggc ccgtatcccc cagggtgtctg caggctcaaa 1680
gagcagcgag aagcgttcag aggaaagcga tcccgtgcc ccttccccgt gcccgggctg 1740
tccccgcacg ctgcccggctc ggggatgcgg ggggagcgcc ggaccggagc ggagccccgg 1800
gcggctcgct gctgccccct agcgggggag ggacgttaatt acatccctgg gggctttggg 1860
ggggggctgt ccccgtagag ggatccgctg ggtgcaggga attcgtaatc atgttgaggg 1920
ctgtttcctg tgtgaaattg ttatccgctc acaattccac acaacatacg agccggaagc 1980
ataaagtgtg aagcctgggg tgcctaatag gtgagctaac tcacattaat tgcgttgctc 2040
tactgcccgc ctttccagtc gggaaacctg tgcgtgccagc tgcattaatg aatcggccaa 2100
cgcgcgggga gaggcggttt gcgtattggg cgctcttcog cttcctcgct cactgactcg 2160
ctgcgctcgg tcttccggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg 2220
ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag 2280
gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg cccccctgac 2340
gagcatcaca aaaatcgacg ctcaagttag aagctccctc tctcccttcg ggaagcgtgg 2400
taccaggcgt tccccctgg aagctccctc cgctcttcog cgtcttctca tagctcacgc 2460
accggatacc tgtccgcctt gtaggtcggt gtaggtcggt cgctccaagc tgggctgtgt 2520
tgtaggtatc tcagttcggg cgccttatcc ggcagcagcc actggttaaca ggattagcag agcgaggtat 2580
cccgttcagc ccgaccgctg ggcagcagcc cttgaagtgg gttaccttcg gaaaaagagt tggtagctct 2640
agacacgact tatcgccact ggcagcagcc gctgagcca ggtggttttt ttgtttgcaa gcagcagatt 2700
gtaggcggtg ctacagagtt cttgaagtgg gctgaagcca cgtggttagc cctttgatct tttctacggg 2760
gtatttggtg tctgcgctct cgctggttag tcaagaagat ttggtcatga gattatcaaa aaggtacttc 2820
tgatccggca aacaaaccac cgtggttagc tgaagaagat cttggtcatga gattatcaaa aaggtacttc 2880
acgcgcagaa aaaaaggatc ttaagggatt aaaaatgaag atgttataat cttactcccc 2940
cagtggaaac aaaactcacg ttaagggatt aaaaatgaag atgttataat cttactcccc 3000
acctagatcc ttttaaatga acttggtctg acagttacca ctgactcccc tgcaatgata 3060
tttgcgttat ccatagttgc tgcaatgata agccggaagg taattgttgc tgccattgct 3120
ttaccatctg gcccagtgcc agccggaagg taattgttgc tgccattgct 3180
ttatcagcaa taaccagcc taattgttgc tgccattgct 3240
tccgcctcca tccagttat tgccattgct cggttcccaa cctcttcggt 3300
aatagtttgc gcaacgttgt tgccattgct cggttcccaa cctcttcggt 3360
ggtatggctt cattcagctc cggttcccaa cctcttcggt 3420
ttgtgcaaaa aagcgggttag tatggcagca tgggtgagta 3480
gcagtgttat cactcatggt tgggtgagta cccggcgta 3540
gtaagatgct tttctgtgac tgggtgagta cccggcgta 3600
cggcgaccga gttgctcttg cccggcgta cccggcgta 3660
actttaaaag tgctcatcat tggaaaacgt gatgtaaccc 3720
ccgctgttga gatccagttc tgggtgagca atggtgaata 3780
tttacttttc ccagcgtttc atggtgaata tctcatgagc 3840
ggaataaggg cgacacggaa atggtgaata tctcatgagc 3900
agcatttatc aggggttattg tctcatgagc cacatttccc 3960
aaacaaatag ggggtccgcg ctttattacc tccgtgcggg 4020
gcaactgttg ggaagggcga ttaagtggg taatagcact 4080
ggggatgtgc tgcaaggcga taatagcact cattgatgag 4140
gtaaaacgac ggccagtcgg taatagcact cattgatgag 4200
tatacagaca tgataagata aatttgtgat gctattgctt 4260
aaaaaatgct ttatttgtga aatttgtgat gctattgctt 4320
tgcaataaac aagttggggg gggcgaaaga aaacgattcc 4380
catccagccg gcgtcccga cgtgtcagtc ctgctcctcg 4440
ggaatcgaaa tctcgtagca cgtgtcagtc ctgctcctcg 4500

```

<210> 118

<211> 17384

<212> DNA

<213> Artificial Sequence

<220>

<223> pFK161 Plasmid

<400> 118

gcgacacgagg	gagcttccag	ggggaaacgc	ctgggtatctt	tatagtcctg	tcgggggtttc	60
gccacctctg	acttgagcgt	cgatTTTTgt	gatgctcgtc	aggggggagg	agcctatgga	120
aaaacgccag	caacgcggcc	TTTTtacggt	tcctggcctt	ttgctggcct	tttgctcaca	180
tggtctttcc	tgcgttatcc	cctgattctg	tggataaccg	tattaccgcc	tttgagttag	240
ctgataccgc	tcgccgcagc	cgaacgaccg	agcgcagcga	gtcagtgagc	gaggaagcgg	300
aagagcgctg	acttccgcgt	ttccagactt	tacgaaacac	ggaaaccgaa	gaccattcat	360
gttggttctc	aggtcgcaga	cgttttgcag	cagcagtcgc	ttcacgttcg	ctcgcgctac	420
ggtgattcat	tctgctaacc	agtaaggcaa	ccccgccagc	ctagccgggt	cctcaacgac	480
aggagcacga	tcatgcgcac	ccgtcagatc	cagacatgat	aagatacatt	gatgagtttg	540
gacaaaccac	aactagaatg	cagtgaaaaa	aatgctttat	ttgtgaaatt	tgtgatgcta	600
ttgctttatt	tgtaaccatt	ataagctgca	ataaacaagt	taacaacaac	aattgcattc	660
attttatggt	tcagggttcag	ggggagggtg	gggaggtttt	ttaaagcaag	taaaacctct	720
acaaatgtgg	tatggctgat	tatgatctct	agtcaaggca	ctatacatca	aatattcctt	780
attaaccctc	ttacaaatta	aaaagctaaa	ggtacacaat	ttttgagcat	agttattaat	840
agcagacact	ctatgcctgt	gtggagtaag	aaaaaacagt	atgttatgat	tataactggt	900
atgcctactt	ataaagggtta	cagaatattt	ttccataatt	ttcttgata	gcagtgagc	960
tttttccttt	gtggtgtaaa	tagcaaagca	agcaagagtt	ctattactaa	acacagcatg	1020
actcaaaaaa	cttagcaatt	ctgaaggaaa	gtccttgggg	tcctctacct	ttctctctct	1080
ttttggagga	gtagaatggt	gagagtcagc	agtagcctca	tcactactag	atggcatttc	1140
ttctgagcaa	aacaggtttt	cctcattaaa	ggcattccac	cactgctccc	attcatcagt	1200
ttcataggtt	ggaatctaaa	atacacaac	aattagaatc	agtagtttaa	cacattatac	1260
acttaaaaat	tttatattta	ccttagagct	ttaaatctct	gtaggtagtt	tgtccaatta	1320
tgtcacacca	cagaagtaag	gttccttcac	aaagatccgg	accaaagcgg	ccatcgtgcc	1380
ttcccactcc	tgtagttcgg	gggcatggat	gcgcgtagat	ccgctgctgg	tttcttggat	1440
gccgacggat	ttgcactgcc	ggtagaactc	gcgaggtcgt	ccagcctcag	gcagcagctg	1500
aaccaactcg	cgaggggatc	gagcccgggg	tgggcgaaga	actccagcat	gagatccccg	1560
cgctggagga	tcattccagcc	ggcgtcccgg	aaaacgattc	cgaagcccaa	cctttcatag	1620
aaggcggcgg	tggaaatcgaa	atctcgtgat	ggcaggttgg	gcgtcgcctg	gtcggtcatt	1680
tcgaacccca	gagtcccgtc	cagaagaact	cgtcaagaag	gcgatagaag	gcgatgcgct	1740
gcgaatcggg	agcggcgata	ccgtaaagca	cgaggaagcg	gtcagcccat	tcgcgcgcaa	1800
gctcttcagg	aatatcacgg	gtagccaaag	ctatgtcctg	atagcgggtc	gccacaccca	1860
gccggccaca	gtcgatgaat	ccagaaaagc	ggccattttc	caccatgata	ttcggcaagc	1920
aggcatcgcc	atgggtcacg	acgagatcct	cgccgtcggg	atgcgcgcct	tgagcctggc	1980
gaacagttcg	gctggcgcgga	gcccctgatg	ctcttctgct	agatcatcct	gatcgacaag	2040
accggcttcc	atccgagtac	gtgctcgctc	tatgcgatgt	ttcgcttggg	ggtcgaatgg	2100
gcaggtagcc	ggatcaagcg	tatgcagccg	ccgcattgca	tcagccatga	tggatacttt	2160
ctcggcagga	gcaaggtagg	atgacaggag	atcctgcccc	ggcacttcgc	ccaatagcag	2220
ccagtcctct	cccgcttcag	tgacaacgct	gagcacagct	gcgcaaggaa	cgcccgctcg	2280
ggccagccac	gatagcccg	ctgcctcgct	ctgcagttca	ttcagggcac	cggacaggta	2340
ggtcttgaca	aaaagaaccg	ggcgccccct	cgctgacagc	cggaacacgg	cggcatcaga	2400
gcagccgatt	gtctgtttgt	cccagtcata	gccgaatagc	ctctccaccc	aagcggccgg	2460
agaacctgcg	tgcaatccat	cttggtcaat	catgcgaaac	gatcctcatc	ctgtctcttg	2520
atcagatctt	gatccccctg	gccatcagat	ccttggcggc	aagaaaagcca	tccagtttac	2580
tttgcaggcg	ttcccaacct	taccagaggg	cgcccgagct	ggcaattccg	gttcgcttgc	2640
tgctccataa	accgcccagc	ctagctatcg	ccatgtaagc	ccactgcaag	ctacctgctt	2700
tctcttttgc	ccttgctttt	cccttgtcca	gatagcccag	tagctgacat	tcattccggg	2760
tcagcacctg	ttctgctggg	tggtctttct	cgtgttccgc	ttccttttag	agcccttgcg	2820
ccctgagtgc	ttgcggcgac	gtgaaagctt	tttgcaaaag	cctaggcctc	caaaaaagcc	2880
tcctcactac	ttctggaata	gctcagaggg	cgaggcgccc	taaataaaaa	aaattagtca	2940
gccatggggc	ggagaatggg	cggaaactgg	cgaggttagg	ggcgggatgg	gcggagttag	3000
gggcgggact	atggttgctg	actaattgag	atgcatgctt	tgcatacttc	tgctgctggg	3060
ggagcctggg	gactttccac	acctgggttg	tgactaattg	agatgcatgc	tttgcatact	3120
tctgcctgct	ggggagcctg	gggaactttc	acaccctaac	tgacacacat	tccacagccg	3180
gatctgcagg	acccaacgct	gcccagagatg	cgccgcgtgc	ggctgctgga	gatggcggac	3240
gcgatggata	tggtctgcca	agggttggtt	tgcgcaattc	cagttctccg	caagaattga	3300
ttggctccaa	ttcttgaggt	ggtgaatccg	ttagcgaggt	gccgcccggc	tccattcagg	3360
tcgaggtggc	ccggctccat	gcaccgcgac	gcaacgcggg	gaggcagaca	aggtataggg	3420
cggcgccctc	aatccatgcc	aaccogttcc	atgtgctcgc	cgaggcgcat	aaatcgccgt	3480
gacgatcagc	ggtccaatga	tcgaagttag	gttggttaaga	gccgcgagcg	atccttgaag	3540
ctgtccctga	tggtcgatcat	ctacctgctt	ggacagcatg	gcctgcaacg	cggcatcccg	3600
atgccgcccg	aagcgagaag	aatcataatg	gggaaggcca	tccagcctcg	cgtcgcgaa	3660
gccagcaaga	cgtagcccag	cgcgtcgggc	cgccatgccc	gcgataatgg	cctgcttctc	3720
gccgaaacgt	ttggtggcgg	gaccagtgc	gaaggcttga	gcgagggcgt	gcaagattcc	3780
gaataccgca	agcgacaggg	cgatcatcgt	cgcgctccag	cgaaagcggt	cctcgccgaa	3840
aatgaccag	agcgctgccg	gcacctgtcc	tacgagttgc	atgataaaga	agacagtcac	3900
aagtgcggcg	acgatagtca	tgccccgcgc	ccaccggaag	gagctgactg	ggttgaaggc	3960
tctcaagggc	atcggtcgac	gctctccctt	atgcgactcc	tgcattagga	agcagccag	4020

tagtagggttg	aggccggttga	gcaccgcccgc	cgcaagggaat	ggtgcatgca	aggagatggc	4080
gccccacagt	cccccgccca	cgggccctgcc	accataccca	cgccgaaaca	agcgctcatg	4140
agcccgaaagt	ggcgagcccg	atcttcccca	tccgtgatgt	cgccgatata	ggcgccagca	4200
accgcacctg	tggcgccggt	gatgccggcc	acgatgcgtc	cgccgtagag	gatcttggca	4260
gtcacagcat	gcgcataatcc	atgcttcgac	catgcgctca	caaagtaggt	gaatgcgcaa	4320
tgtagtacc	acatcgatcat	cgttttccac	tgctctcgcg	aataaagatg	gaaaatcaat	4380
ctcatggtaa	tagtccatga	aaatccttgt	attcataaat	cctccaggta	gctatatgca	4440
aattgaaaca	aaagagatgg	tgatctttct	aagagatgat	ggaatctccc	ttcagtatcc	4500
cgatgggtcaa	tgcgctggat	atgggataga	tgggaatatg	ctgattttta	tgggacagag	4560
ttgcgaactg	ttcccaacta	aaatcatttt	gcacgatcag	cgactacga	actttaccca	4620
caaatagtca	ggtaatgaat	cctgatataa	agacagggtg	ataaatcagt	cttctacgcg	4680
catcgcacgc	gcacaccgta	gaaagtcttt	cagttgtgag	cctgggcaaa	ccgttaactt	4740
tccggcgctt	tgctgtgcga	caggctcacg	tctaaaagga	aataaatcat	gggtcataaa	4800
attatcacgt	tgtccggcgc	ggcgacggat	gttctgtatg	cgctgttttt	ccgtggcgcg	4860
ttgctgtctg	gtgatctgcc	ttctaaatct	ggcacagccg	aattgcgcga	gcttgggttt	4920
gctgaaacca	gacacacagc	aactgaatac	cagaaagaaa	atcactttac	ctttctgaca	4980
tcagaaggcg	agaaatttgc	cgttgaacac	ctgggtcaata	cgctgttttg	tgagcagcaa	5040
tattgcgctt	cgatgacgct	tggcggtgag	attgatacct	ctgctgcaca	aaaggcaatc	5100
gacgagctgg	accagcgcat	tccgtgacacc	gtctccttcg	aacttattcg	caatggagtg	5160
tcattcatca	aggacgccgc	tatcgcaaat	gggtgctatcc	acgcagccgc	aatcgaaaca	5220
cctcagccgc	tgaccaatat	ctacaacatc	atccttggta	tccagcgtga	tgagccagcg	5280
cagaacaagg	taacgcgtcag	tgcgcgataag	tctaaaagta	aacctgggtg	tgataccaac	5340
attgaaacgt	tgatcgaaaa	cgcgctgaaa	aacgctgctg	aatgtgcggc	gctggatgtc	5400
acaaagcaaa	tggcagcaga	caagaaagcg	atggatgaac	tggcttccta	tgtccgcacg	5460
gccatcatga	tggaaatgtt	ccccgggtgt	gttatctggc	agcagtgcgg	tcgatagtat	5520
gcaattgata	attattatca	tttgcgggtc	ctttccggcg	atccgccttg	ttacggggcg	5580
gcgacctcgc	gggttttccg	tatttatgaa	aattttccgg	tttaaggcgt	ttccgttctt	5640
cttcgctcata	acttaatgtt	tttattttaa	ataccctctg	aaaagaaagg	aaacgacagg	5700
tgctgaaagc	gagctttttg	gcctctgtcg	tttcctttct	ctgtttttgt	ccgtggaatg	5760
aacaatggaa	gtcaacaaaa	agcagctggc	tgacattttc	ggtgcgagta	tccgtaccat	5820
tcagaactgg	caggaaacagg	gaatgcccgt	tctgcgaggc	ggtggcaagg	gtaatgaggt	5880
gctttatgac	tctgccgcgc	tcataaaatg	gtatgccgaa	agggatgctg	aaattgagaa	5940
cgaaaagctg	cgcggggagg	ttgaagaact	gcggcaggcc	agcgaggcag	atccacagga	6000
cgggtgtggg	cgccatgatc	gcgtagtoga	tagtggctcc	aagttagcgaa	gcgagcagga	6060
ctgggcggcg	gcaaagcggg	cggacagtcg	tccgagaacg	ggtgcgcata	gaaattgcac	6120
caacgcata	agcgctagca	gcacgccata	gtgactggcg	atgctgtcgg	aatggacgat	6180
atcccgcaag	aggccccgca	gtaccggcat	aaccaagcct	atgcctacag	catccagggt	6240
gacgggtgcc	aggatgacga	tgagcgcatt	gttagatttc	atacacggtg	cctgactgcg	6300
ttagcaattt	aactgtgata	aactaccgca	ttaaagctta	tcgatgataa	gcgggtcaaac	6360
atgagaattc	gcgcccgctc	ttctcgttct	ggccagggcg	cctcgctctc	ccaccaccac	6420
cgtctgcggg	tgggtgtgtg	aaggcagggg	tgcggctctc	cggcccgacg	ctgcccccg	6480
cgcacttttc	tcagtgggtc	gcgtgggtct	tgtggatgtg	tgaggcgccc	ggttgtgccc	6540
tcacgtgttt	cactttgggtc	gtgtctcgct	tgaccatgtt	cccagagtcg	gtggatgtgg	6600
cgggtggcgt	tgcataccct	tcccgtctgg	tgtgtgcacg	cgctgtttct	tgtaagcgtc	6660
gaggtgtctc	tggagcgttc	caggtttgtc	tcttaggtgc	ctgcttctga	gctggtgtgc	6720
gcgctcccca	ttccctgggtg	tgcctccggt	gctccgtctg	gctgtgtgcc	ttcccggttg	6780
tgtctgagaa	gcccggtgaga	gggggggtoga	ggagagaagg	aggggcaaga	cccccttctc	6840
tcgtcgggtg	aggcgccccc	cccgcgacta	gtacgcctgt	gcgtagggct	ggtgctgagc	6900
ggtcgcggct	ggggttggaa	agttttctcga	gagactcatt	gctttcccgt	ggggagcttt	6960
gagaggcctg	gctttcgggg	gggaccgggt	gcagggtctc	cctgtccgcg	ggatgctcag	7020
aatgcccttg	gaagagaacc	ttcctgttgc	cgcagacccc	cccgcgcggg	cgcccgcggtg	7080
ttggtcttct	ggtttccctg	tgtgctcgct	gcacgcaccc	tctctcggtg	gccggggctc	7140
gtcgggggtt	tgggtccgtc	ccgcccctcag	tgagaaagtt	tccttctcta	gctatcttcc	7200
ggaaaggggt	cgggcttctt	acggtctctc	ggggctctct	ccgaatgggt	ccctggaggg	7260
ctcgccccct	gaccgcctcc	cgcgcgcgca	gcgtttgtct	tctcgtctac	cgcggcccg	7320
ggcctccccg	ctccgagttc	ggggagggat	cacgcggggc	agagcctgtc	tgtcgtcctg	7380
ccgttgctgc	ggagcatgtg	gctcggcttg	tgtgggtggg	ggctggggag	agggctccgt	7440
gcacaccccc	gcgtgcgcgt	actttccctc	cctcctgagg	gccgccgtgc	ggacgggggtg	7500
tgggtaggcg	tgggtaggcg	accgggtccc	caccgctctt	cccgctgctc	accgctgctc	7560
tccgtcgcgt	gcgtccctct	cgtcgcgctc	cacgactttg	gccgctcccc	cgacggcggc	7620
ctgcgcgcgc	cgtgggtgcgt	gctgtgtgct	tctcgggctg	tgtgggtgtg	tcgcctcgcc	7680
cccccttctc	cgcggcagcg	ttcccacggc	tggcgaaatc	gcgggagtc	tccttccccct	7740
cctcgggggt	gagaggggtc	gtgtctggcg	ttgattgac	tcgctctcgg	ggacggggacc	7800
gttctgtggg	agaaacggctg	ttggccggcg	cgggcgcgac	gtcggacgtg	gggacccact	7860
gccgctcggg	ggtcttctgc	ggtaggcac	ggtgtgtcgg	catcggtctc	tctctcgtgt	7920
cgggtgtcgc	tcctcgggct	cccggggggc	cgtcgtgttt	cgggtcggct	cggcgctgca	7980
ggtgtgggtg	gactgctcag	gggagtggtg	cagtgtgatt	cccgccgggt	ttgcctcgcg	8040

-86-

tgccttgacc	gggtccgacgc	ccgagcgggtc	tctcggtccc	ttgtgaggac	ccccctccgg	8100
gagggggccc	tttcggccgc	ccttgccgtc	gtcgccggcc	ctcgttctgc	tgtgtcggtc	8160
ccccctccc	gctcgccgca	gccggtcttt	ttcctctct	ccccccctct	cctctgactg	8220
accggtggcc	gtgctgtcgg	accccccgca	tgggggcggc	cgggcacgta	cgcgtccggg	8280
cggtcacccg	ggctctgggg	gggggcccag	gggtaagaaa	gtcggctcgg	cgggcggggag	8340
gagctgtggt	ttggagggcg	tccccggccc	gcgcccggtg	cgggtgtctg	cgcgggtctg	8400
gagagggctg	cgtgcgaggg	gaaaagggtg	ccccgcgagg	gcaaagggaa	agaggctagc	8460
agtggtcatt	gtcccgcagg	tgtggtggtc	tgttggccga	gggtgcgtctg	gggggctcgt	8520
ccggccctgt	cgtccgtcgg	gaaggcgcgt	gttggggcct	gccggagtgc	cgaggtgggt	8580
accctggcgg	tgggattaac	cccgcgcgcg	tgtcccgggtg	tggcgggtggg	ggctccggtc	8640
gatgtctacc	tccctctccc	cgaggtctca	ggccttctcc	gcgcgggctc	tcggcccttc	8700
cctcgttcct	ccctctcgcc	gggttcaagt	cgtcgtcgca	cctcccctcc	tccgtccttc	8760
catctctcgc	gcaatggcgc	cgcccagagt	cacgggtgggt	tcgtcctccg	cctccgcttc	8820
tgcgcggggg	ctggccgctg	tccggtctct	cctgcccgac	cccgttggc	gtggtcttct	8880
ctcgccgggt	tgcgggactc	ctggcttcgc	ccggagggtc	agggggcttc	ccggttcccc	8940
gacgttgccg	ctcgctgctg	tgtgcttggg	gggggcccgc	tgcggcctcc	gccgcgccgt	9000
gagccccctg	cgcacccgcc	gggtgtcgggt	ttcgccgcgc	ggtcagttgg	gccctggcgt	9060
tgtgtcgcgt	cgggagcgtg	tccgcctcgc	ggcggtctaga	cgcgggtgtc	gccgggctcc	9120
gacgggtggc	ctatccaggg	ctcgcccccg	ccgacccccg	cctgcccgtc	ccggtgggtg	9180
tcggttgggt	ggggagtgaa	tgggtgtacc	gggtcattccc	tcccgcgtgg	tttgactgtc	9240
tgcgcgggtg	cgcgcttctc	tttcgcgcaa	ccccacgcc	aaccacccac	cctgctctcc	9300
cggcccggtg	cggtcgacgt	tccggctctc	cagatgccga	gggggttcggg	atttgtgccc	9360
gggacggagg	ggagagcggg	taagagaggt	gtcggagagc	tgtcccgggg	cgacgctcgg	9420
gttggctttg	ccgcgtgcgt	gtgctcgcgg	acgggttttg	tccgaacccg	acggggtcgg	9480
tccggccgca	tgcactctcc	cgttcgcgcg	gagcgcgccg	ccggtcacc	cccgttttgt	9540
cctcccgcga	ggctctccgc	cgcgcgcgcg	tctcctcct	ctctcgcgct	ctctgtcccg	9600
cctggctcctg	tcccaccccc	gacgctccgc	tcgcgcttcc	ttacctgggt	gatcctgcca	9660
ggtagcatat	gcttgtctca	aagattaagc	catgcatgtc	taagtacgca	cggccgggtac	9720
agtgaaactg	cgaatggctc	attaaatcag	ttatgggtcc	tttgggtcgt	cgtcctctct	9780
ctacttggat	aactgtggta	attctagagc	taatacatgc	cgcggggcgc	tgacccccct	9840
tcccgggggg	ggatgcgtgc	atttatcaga	tcaaaaaccaa	cccgggtgagc	tccctcccgg	9900
ctccggcccg	gggtcgggcg	ccggcggctt	ggtgactcta	gataacctcg	ggccgatcgc	9960
acgccccccg	tggcggcgac	gacccattcg	aacgtctgcc	ctatcaactt	tcgatggtag	10020
tcgcccgtgc	taccatgggtg	accacgggtg	acggggaatc	aggggttcgat	tccggagagg	10080
gagcctgaga	aacggctacc	acatccaagg	aaggcagcag	gcgcgcaaat	taccactcc	10140
cgaaccgggg	aggtagtgac	aaaaaataac	aatacaggac	tcttccgagg	ccctgtaatt	10200
ggaatgagtc	cactttaaat	cctttaacga	ggatccattg	gagggcaagt	ctggtgccag	10260
cagccgcggg	aattccagct	ccaatagcgt	atattaaagt	tgttcgagtt	aaaaagctcg	10320
tagttggatc	ttgggagcgg	gcgggcggtc	cgccgcgagg	caggtcaccg	cccgtccccg	10380
ccccttgctc	ctcgcgccc	cctcgtatgc	cttagctgag	tgtcccgcgg	ggcccgaagc	10440
gtttactttg	aaaaaattag	agtgttcaaa	gcaggcccga	gccgcctgga	taccgcagct	10500
aggaataatg	gaataggacc	gcggttctat	tttgttgggt	ttcggaactg	aggccatgat	10560
taagaggggac	ggccgggggg	attcgtattg	cgccgctaga	ggtgaaatcc	ttggaccggc	10620
gcaagacgga	ccagagcgaa	agcatttgcc	aagaatgttt	tcattaatca	agaacgaaag	10680
tcggaggttc	gaagacgatc	gcagaccgtc	tgtgttccga	ccataaacga	tgcccatagg	10740
cgatgcggcg	gcgttattcc	catgacccgc	cgggcagctt	ccgggaaacc	aaagtctttg	10800
ggttccgggg	ggagtatggt	tgcaaagctg	aaacttaaaag	gaattgacgg	aagggcacca	10860
ccaggagtg	gcctgcgggt	taatttgact	caacacggga	aacctcacc	ggcccgggca	10920
cggacaggat	tgacagattg	atagctcttt	ctcgattccg	tgggtgggtg	tgcatggccg	10980
ttcttagttg	gtggagcgat	ttgtctggtt	aattccgata	acgaacgaga	ctctggcatg	11040
ctaactagtt	acgcgacccc	cgagcggctc	gcgtccccc	acttcttaga	gggacaagtg	11100
gcgttcagcc	acccgagatt	gagcaataac	aggtctgtga	tgcccttaga	tgtccggggc	11160
tgcacgcgcg	ctacactgac	tggctcagcg	tgtgctacc	ctgcgcgggc	aggcgcgggt	11220
aaccggttga	accacattcg	tgatggggat	cggggattgc	aattattccc	catgaacgag	11280
gaattcccag	taagtgcggg	tcataagctt	gcgttgatta	agtcocctgcc	ccttgtagac	11340
accgcccgtc	gctactaccg	attggatggg	ttagtgaggc	cctcggatcg	gccccgcggg	11400
ggtcggccca	cggccctggc	ggagcgcgtg	gaagacggtc	gaacttgact	atctagagga	11460
agtaaaaagt	gtaacaagg	ttccgtaggt	gaacctgcgg	aaggatcatt	aaacggggag	11520
ctgtggaggga	gcggcgcggt	cgccgcctct	cccgtcttg	tgtgtgtcct	cgccgggagg	11580
cgcgtgcgtc	ccgggtcccg	tcgcccgcgt	gtggagcggg	gtgtctggag	tgaggtgaga	11640
gaaggggtgg	gtggggctcg	tctgggtccg	tctgggaccg	cctccgattt	ccccctcccc	11700
tcccctctcc	ctcgtccggc	tctgacctcg	ccacctacc	gcggcggcgg	ctgctcgcgg	11760
gcgtcttgcc	tctttcccg	ccggtctctc	cgtgtctacg	aggggcggga	cgtcgttacg	11820
ggtttttgac	ccgtccgggg	ggcgttcggg	cgtcggggcg	cgcgctttgc	tctccgggca	11880
cccatccccg	ccgcggctct	ggcttttcta	cgttggctgg	ggcgggtgtc	gcgtgtgggg	11940
ggatgtgagt	gtcgcgtgtg	ggctcgcgg	tcccgatgcc	acgcttttct	ggcctcgcgt	12000
gtcctccccg	ctcctgtccc	gggtacctag	ctgtcgcgtt	ccggcgcgga	ggtttaagga	12060

-87-

ccccgggggg	gtcgccctgc	cgcccccagg	gtcggggggg	gggtggggccc	gtaggggaagt	12120
cggtcgttcg	ggcggtctctc	cctcagactc	catgaccctc	ctccccccgc	tgccgcggtt	12180
cccgagggcg	cggtcgtgtg	gggggggtgga	tgtctggagc	ccccctcgggc	gccgtggggg	12240
cccgacccgc	gccgcgcggt	tgcccgattt	ccgcgggtcg	gtcctgtcgg	tgccggtcgt	12300
gggttccccg	gtcgttcccc	tgttttttcg	ctcccgaccc	tttttttttc	ctccccccca	12360
cacgtgtctc	gtttcgtttc	tgctggccgg	cctgaggcta	ccccctcggtc	catctgttct	12420
ctctctctctc	cggggagagg	agggcggtgg	tcgttggggg	actgtgcccgt	cgtcagcacc	12480
cgtgagttcg	ctcacacccg	aaataccgat	acgactctta	gcgggtggatc	actcggctcg	12540
tgctcgatg	aagaacgcag	ctagctgcga	gaattaatgt	gaattgcagg	acacattgat	12600
catcgacact	tcgaacgcac	ttgcggcccc	gggttccctcc	cgggggctacg	cctgtctgag	12660
cgtcggttga	cgatcaatcg	cgtcaccgcg	tgccgtgggt	gctgcgcggc	tgggagtttg	12720
ctcgacgggc	caacccccca	accgggtcgc	ggccctccgt	ctcccgaaagt	tcagacgtgt	12780
gggcggttgt	cggtgtggcg	cgcgcgcgcc	cgtcgcggag	cctgggtctcc	cccgcgcac	12840
cgcgctcgcg	gcttcttccc	gctccgcccgt	tcccgccctc	gcccgtgcac	cccgggtcctg	12900
gcctcgcgct	ggcgccctccc	ggaccgctgc	ctcaccagtc	tttctcggtc	ccgtgccccg	12960
tgggaaaccca	cgcgcgcgcc	gtggcgcccg	ggggtgggcy	cgtccgcate	tgctctggtc	13020
gaggttggcg	gttgagggtg	tgctgcgccc	gaggtgggtg	tcgggtccct	gcggccggcg	13080
ggttgcggg	gtggcggtcg	acgaggggcg	gtcggtcgcc	tgccgtgggt	gtctgtgtgt	13140
gtttgggtct	tgccgtgggg	gaggcggggt	cgaccgctcg	cgggggtggc	gcggtcgccc	13200
ggcgccgcgc	accctccggc	ttgtgtggag	ggagagcgag	ggcgagaacg	gagagaggtg	13260
gtatccccgg	tgccgttcgc	agggagggtt	tggcgtcccc	cgtccgtccg	tccctccctc	13320
cctcggtggg	cgccttcgcg	cgcacgcgg	ccgttagggg	cggtcggggc	ccgtggcccc	13380
cgtggctctt	cttcgtctcc	gcttctcctt	caccggggcg	gtaccgcctc	cggcgccggc	13440
ccgcgggacg	ccgcggcgct	cgtgcgccga	tcgagtcac	ccccgggtgt	tgcgagttcg	13500
gggagggaga	gggcctcgct	gacccgttgc	gtcccggtt	ccctgggggg	gaccggcgct	13560
ctgtggcgct	tgccgtcccg	gggttgccgt	tgagtaagat	cctccacccc	cgcgcgccc	13620
ccctcccgcc	ggcctctcgg	ggacccccct	agacggttcg	ccggctcgtc	ctcccgctgc	13680
gccgggtgcc	gtctctttcc	cgcgcgcctc	ctcgctctct	tcttcccgcg	gctgggcgcg	13740
tgtccccct	ttctgaccgc	gacctcagat	cagacgtggc	gaccgcgtga	attttaagcat	13800
attagt cagc	ggaggaaaag	aaactaacca	ggattccctc	agtaacggcg	agtgaacagg	13860
gaagagccca	cgcgcgaatc	ccgcgcgcgc	gtcgcgcgct	gggaaatgtg	cgctacggaa	13920
gaccactcc	ccggcgccgc	tcgtgggggg	cccaagtcct	tctgatcgag	gcccagcccc	13980
tgagcggtgt	gaggccggta	gcggcccccg	cgcgcggggc	tcgggtcttc	ccggagtcgg	14040
gttgcttggg	aatgcagccc	aaagcgggtg	gtaaactcca	tctaaggcta	aataccggca	14100
cgagacgat	agtcaacaag	taccgtaagg	gaaagtgtga	aagaactttg	aagagagagt	14160
tcaagagggc	gtgaaaacgt	taagaggtaa	acgggtgggg	tccgcgcagt	ccgcccggag	14220
gattcaaccc	ggcggcgcgc	gtccggccgt	gcccggtggg	cccggcggtt	ctttcccgct	14280
ccccgttcc	ccgcacccct	ccaccgcgc	gtcgttcccc	tcttccctcc	cgcgtccggc	14340
gcctccggcg	gcgggcgcgg	ggggtgggtg	ggtgggtggc	cgcgggcggg	gcccgggggtg	14400
gggtcggcgg	gggcgcgcc	ccggccggcg	accggccgcc	gcccggcgca	cttccaccgt	14460
ggcggtgcgc	cgcgacccgc	tcggggacgg	ccgggaaggc	ccggtgggga	aggtggctcg	14520
ggggggggcg	cgcgtctcag	ggcgcgccga	accactcac	cccagtggtt	acagccctcc	14580
ggccgcgctt	tcgcgcgaatc	ccggggccga	ggaagccaga	taccgcctgc	cgcgtctctc	14640
ctctcccccc	gtccgcctcc	cgggcggggc	tggggtggg	ggccgggcgc	ccccctccc	14700
ggcgcgacgc	ctctccacc	cccctccgtc	gcctctcgc	gggcccgggtg	gggggcgggg	14760
cggactgtcc	ccagtgcgcc	ccgggcgtcg	tcgcgcgcgc	gggtcccggg	gggaccgtcg	14820
gtcacgcgtc	tcccgacgaa	gccgagcgca	cggggtcggc	ggcgatgtcg	gctaccacc	14880
cgacccgtct	tgaaacacgg	accaaggagt	ctaaccgcgtg	cgcgagtcag	gggctcgtcc	14940
gaaagccgcg	gtggcgcaat	gaagggtgaag	ggcccggccc	ggggggccga	ggtgggatcc	15000
cgaggccctc	ccagtccgcc	gagggcgcac	caccggcccg	tctcgccgcg	cgcgcggggg	15060
aggtggagca	cgagcgtacg	cgttaggacc	cgaaagatgg	tgaactatgc	ttgggcaggg	15120
cgaagccaga	ggaaactctg	gtggaggtcc	gtagcggtcc	tgacgtgcaa	atcggtcgtc	15180
cgacctgggt	atagggggcg	aagactaatc	gaaccatcta	gtagctgggt	ccctccgaag	15240
ttccctcag	gatagctggc	gctctcgtc	ccgacgtacg	cagttttatc	cggtaaacgc	15300
aatgattaga	ggtccttggg	ccgaaacgat	ctcaacctat	tctcaaacct	taaatgggta	15360
agaagcccg	ctcgctggcg	tggagccggg	cgtggaatgc	gagtgcctag	tgggccactt	15420
ttggtaagca	gaactggcgc	tgccggatga	accgaacgcc	gggttaaggc	gcccgatgcc	15480
gacgctcatc	agacccccaga	aaagggtgtg	gttgatatag	acagcaggac	ggtggccatg	15540
gaagtccgaa	tcgcctaagg	agtgtgtaac	ccgaatcaac	ccgaatcaac	tagccctgaa	15600
aatggatggc	gctggagcgt	cgggcccata	cccggccgtc	gccgcagtcg	gaacggaacg	15660
ggacggggagc	ggcgcgcaat	tcttgaagac	gaaaggccct	cgtgatacgc	ctatttttat	15720
aggttaatgt	catgataata	atggtttctt	agacgtcagg	tggcactttt	cggggaaatg	15780
tgcgcggaac	ccctatttgt	ttattttttt	aaatacattc	aaatatgtat	ccgctcatga	15840
gacaataaac	ctgataaaatg	cttcaataat	attgaaaaag	gaagagtatg	agtattcaac	15900
atttccgtgt	cgcctttatt	cccttttttt	cggcattttg	cttccgtgtt	ttgctcacc	15960
agaaacgcgt	gtgaaagtaa	aagatgctga	agatcagttg	gggtgcacgag	tgggttacat	16020
cgaactggat	ctcaacacgc	gtaagatcct	tgagagtttt	cgcgccgaag	aacgtttttc	16080

-88-

aatgatgagc	actttttaag	ttctgctatg	tggcgcggtg	ttatcccgtg	ttgacgcggg	16140
gcaagagcaa	ctcggtcgcc	gcatacacta	ttctcagaat	gacttggttg	agtactcacc	16200
agtcacagaa	aagcatctta	cggatggcat	gacagtaaga	gaattatgca	gtgctgccat	16260
aaccatgagt	gataacactg	cggccaactt	acttctgaca	acgatcggag	gaccgaagga	16320
gctaaccgct	tttttgacac	acatggggga	tcagtgaact	cgccttgatc	gttgggaacc	16380
ggagctgaat	gaagccatac	caaacgcagc	gcgtgacacc	acgatgcctg	cagcaatggc	16440
aacaacgttg	cgaaaactat	taactggcga	actacttact	ctagcttccc	ggcaacaatt	16500
aatagactgg	atggaggcgg	ataaagttgc	aggaccactt	ctgcgctcgg	cccttccggc	16560
tggctggttt	attgctgata	aatctggagc	cgggtgagcgt	gggtctcgcg	gtatcattgc	16620
agcactgggg	ccagatggta	agccctcccc	tatcgtagtt	atctacacga	cggggagtca	16680
ggcaactatg	gatgaacgaa	atagacagat	cgtgagata	ggtgcctcac	tgattaagca	16740
ttggttaactg	tcagaccaag	tttactcata	tatacttttag	attgatttaa	aacttcatth	16800
ttaatttaaa	aggatctag	tgaagatcct	ttttgataat	ctcatgacca	aaatccctta	16860
acgtgagttt	tcgttccact	gagcgtcaga	ccccgtagaa	aagatcaaag	gatcttcttg	16920
agatcctttt	tttctgcgcg	taactctgctg	cttgcaaaaca	aaaaaaccac	cgtaccagc	16980
gggtggtttg	ttgcccgatc	aagagctacc	aactcctttt	ccgaaggtaa	ctggcttcag	17040
cagagcgagc	ataccaataa	ctgtccttct	agtgtagccg	tagttaggcc	accacttcaa	17100
gaactctgta	gcaccgccta	catacctcgc	tctgctaata	ctgttaccag	tggtgctgct	17160
cagtggcgat	aagtcgtgtc	ttaccggggt	ggactcaaga	cgatagttac	cggataaggc	17220
gcagcggtcg	ggctgaacgg	gggggttcgtg	cacacagccc	agcttgagag	gaacgacct	17280
caccgaactg	agatacctac	agcgtgagct	atgagaaagc	gccacgcttc	cgaagggaga	17340
aaggcggaca	ggtatccggt	aagcggcagg	gtcggaaacag	gaga		17384

<210> 119

<211> 2814

<212> DNA

<213> Artificial Sequence

<220>

<223> pLITMUS38 Plasmid

<400> 119

gttaactacg	tcagggtggca	cttttcggggg	aaatgtgctg	ggaaccccta	tttgtttatt	60
ttttctaaata	cattcaataa	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	120
ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	180
ttttgctggca	ttttgctctt	ctgtttttgc	tcacccagaa	acgctgggtg	aagtaaaaga	240
tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	acagcggtaa	300
gatccttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagtctt	360
gctatgtggc	gcggtattat	cccgtgttga	cgcggggcaa	gagcaactcg	gtcgcgcgat	420
acactattct	cagaattgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	480
tggcatgaca	gtaagagaat	tatgacagtgc	tgcataaacc	atgagtgata	acactgcggc	540
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaacaacat	600
gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	ccataccaaa	660
cgacgagcgt	ctaccacga	tgctgttagc	aatggcaaca	acgttgcgca	aactattaac	720
tggcgaaact	gtactctag	cttcccgcca	acaattaata	gactggatgg	aggcgataaa	780
agttgcagga	ccacttctgc	gctcgccctt	tccggctggc	tggtttattg	ctgataaatc	840
tggagccggg	gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggtaaagg	900
ctcccgtatc	gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	960
acagatcgct	gagataggtg	cctcactgat	taagcatttg	taactgtcag	accaagttaa	1020
ctcatatata	ctttagattg	atttaccocg	gttgataatc	agaaaagccc	caaaaacagg	1080
aagattgtat	aagcaaatat	ttaaattgta	aacgttaata	ttttgttaaa	attcgcgtta	1140
aatttttggt	aaatcagctc	attttttaac	caataggccg	aaatcggcaa	aatcccttat	1200
aaatcaaaaag	aatagcccga	gatagggttg	agtgttgctc	cagtttgtaa	caagagtcca	1260
ctattaaaga	acgtggactc	caacgtcaaa	gggcgaaaaa	ccgtctatca	ggcgatggc	1320
ccactacgtg	aaccatcacc	caaatacagt	tttttggggg	cgaggtgccg	taaagcacta	1380
aatcggaacc	ctaaaggag	cccccgattt	agagcttgac	ggggaaagcg	aacgtggcga	1440
gaaaggaagg	gaagaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcggcta	1500
cgctgcgcgt	aaccaccaca	cccgcgcgcg	ttaatgcgcg	gctacagggc	gcgtaaaagg	1560
atctaggtga	agatcctttt	tgataatctc	atgaccaaaa	tcccttaacg	tgagttttcg	1620
ttccaactga	cgtcagaccc	cgtagaaaag	atcaaaaggat	cttcttgaga	tccttttttt	1680
ctgcgcgtaa	tctgctgctt	gcaaacaaaa	aaaccaccgc	taccagcggg	ggtttggttg	1740
ccggatcaag	agctaccaac	tctttttccg	aaggtaactg	gcttcagcag	agcgcagata	1800
ccaaataactg	ttcttctagt	gtagccgtag	ttagccacc	acttcaagaa	ctctgtagca	1860
ccgcctacat	acctcgtctt	gctaactcgt	ttaccagtg	ctgctgccag	tggcgataag	1920
tcgtgtctta	ccgggttgga	ctcaagacga	tagttaccgg	ataaggcgca	gcggctcggc	1980
tgaacggggg	gttcgtgcac	acagcccagc	ttggagcgaa	cgacctacac	cgaactgaga	2040
tacctacagc	gtgagctatg	agaaagcgcc	acgcttcccg	aaggggagaa	ggcggacagg	2100

-89-

tatccggtaa	gcggcagggg	cggaaacagga	gagcgcacga	gggagcttcc	aggggggaaac	2160
gcctgggtatc	tttatagtc	tgctggggtt	cgccacctct	gacttgagcg	tcgatttttg	2220
tgatgctcgt	cagggggggc	gagcctatgg	aaaaacgcca	gcaacgcggc	ctttttacgg	2280
ttcctggcct	tttgctggcc	ttttgctcac	atgtaatgtg	agttagctca	ctcattaggc	2340
accccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tgtggaattg	tgagcgggata	2400
acaattttcac	acaggaaaca	gctatgacca	tgattacgcc	aagctacgta	atacgactca	2460
ctagtggggc	ccgtgcaatt	gaagccggct	ggcgccaagc	ttctctgcag	gatatctgga	2520
tcacgaatt	cgctagcttc	ggcctgacg	cgtctccgga	tgtacaggca	tgcgctcgacc	2580
ctctagtcaa	ggccttaagt	gagtcgtatt	acggactggc	cgtcgtttta	caacgctcgtg	2640
actgggaaaa	ccctggcggt	acccaactta	atcgcccttg	agcacatccc	ccttttcgcca	2700
gctggcgtaa	tagcgaagag	gcccgcaccg	atcgcccttc	ccaacagttg	cgcagcctga	2760
atggcgcaatg	gcgcttcgct	tggttaataaa	gcccgccttcg	gcgggctttt	tttt	2814

<210> 120

<211> 2847

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attB Plasmid

<400> 120

gttaactacg	tcagggtggca	cttttcgggg	aaatgtgctg	ggaaccccta	tttgttttatt	60
tttctaaata	cattcaataa	tgtatccgct	catgagacaa	taacctgat	aaatgcttca	120
ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	180
ttttgcggca	ttttgccttc	ctgtttttgc	tcaccagaaa	acgctgggtga	aagtaaaaga	240
tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	acagcggtaa	300
gacctttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagttct	360
gctatgtggc	gcggtattat	cccggtgttg	cgccggggcaa	gagcaactcg	gtcgccgcac	420
acactatttc	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	480
tgccatgaca	gtaagagaat	tatgcagtg	tgccataacc	atgagtata	acactgcggc	540
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaaacat	600
gggggatcat	gtaactcgcc	ttgatcggtg	ggaaccggag	ctgaatgaag	ccataccaaa	660
cgacgagcgt	gacaccacga	tgccctgtag	aatggcaaca	acgttgcgca	aactattaac	720
tggcgaacta	cttactctag	cttcccgcca	acaattaata	gactggatgg	aggcgataaa	780
agttgcagga	ccacttctgc	gctcgccct	tcgggctggc	tggtttattg	ctgataaatc	840
tgagaccggg	gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggttaagg	900
ctcccgtatc	gtagtatatc	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	960
acagatcgct	gagatagggtg	cctcactgat	taagcattgg	taactgtcag	accaagttta	1020
ctcatatata	ctttagattg	atttaccccg	gttgataatc	agaaaagccc	caaaaacagg	1080
aagattgtat	aagcaaata	ttaaattgta	aacgttaata	ttttgttaaa	attcgcggtta	1140
aattttttgtt	aaatcagctc	attttttaac	caataggccg	aaatcggcaa	aatcccttat	1200
aaatcaaaaag	aatagcccga	gataggggtg	agtgttggtc	cagtttgga	caagagtcca	1260
ctattaaaga	acgtggactc	caacgtcaaa	ggcggaaaaa	ccgtctatca	ggcgatggc	1320
ccactacgtg	aaactcacc	tttttgggtt	ctgttggttg	cgaggtgccc	taaaagcata	1380
aatcggaacc	ctaaaggag	cccccgattt	agagcttgac	ggggaaagcg	aacgtggcga	1440
gaaaggaagg	gaagaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcggctca	1500
cgctgcgcgt	aaccaccaca	cccgccgcgc	ttaatgcgcc	gctacagggc	gcgtaaaagg	1560
atctagggtga	agatcctttt	tgataatctc	atgacaaaaa	tcctttaacg	tgagttttcg	1620
ttccactgag	cgtcagacc	cgtagaaaag	atcaaaagat	cttcttgaga	tccttttttt	1680
ctgcgcgtaa	tctgctgctt	gcaaacaaaa	aaaccaccgc	taccagcggg	ggtttggttg	1740
ccggatcaag	agctaccaac	tctttttccg	aaggtaactg	gcttcagcag	agcgagata	1800
ccaaatactg	ttcttctagt	gtagccgtag	ttaggccacc	acttcaagaa	ctctgtagca	1860
ccgctacat	acctcgctct	gctaatoctg	ttaccagtgg	ctgctgccag	tgccgataag	1920
tcgtgtctta	ccgggttgga	ctcaagacga	tagttaccgg	ataaggcgca	gcggctcgggc	1980
tgaacggggg	gttcgtgcac	acagcccagc	ttggagcgaa	cgacctacac	cgaactgaga	2040
tacctacagc	gtgagctatg	agaaagcgcc	acgcttcccg	aaggggagaaa	ggcggacagg	2100
tatccggtaa	gcggcagggg	cggaaacagga	gagcgcacga	gggagcttcc	aggggggaaac	2160
gcctgggtatc	tttatagtc	tgctggggtt	cgccacctct	gacttgagcg	tcgatttttg	2220
tgatgctcgt	cagggggggc	gagcctatgg	aaaaacgcca	gcaacgcggc	ctttttacgg	2280
ttcctggcct	tttgctggcc	ttttgctcac	atgtaatgtg	agttagctca	ctcattaggc	2340
accccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tgtggaattg	tgagcgggata	2400
acaattttcac	acaggaaaca	gctatgacca	tgattacgcc	aagctacgta	atacgactca	2460
ctagtggggc	ccgtgcaatt	gaagccggct	ggcgccaagc	ttctctgcag	gattgaagcc	2520
tgctttttta	tactaacttg	agcgaaatct	ggatccacga	attcgctagc	ttcgcccggtg	2580
acgcgtctcc	ggatgtacag	gcatgcgtcg	accctctagt	caaggcctta	agttagtcgt	2640
attacggact	ggcgttcgct	ttacaacgtc	gtgactggga	aaacctgggc	gttaccacaac	2700

-90-

ttaatcgccct	tgcagcacat	ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	2760
ccgatcgccc	ttcccaacag	ttgcgcagcc	tgaatggcga	atggcgcttc	gcttggtaat	2820
aaagcccgct	tcggcgggct	ttttttt				2847

<210> 121

<211> 4223

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attBBSRpolyA2 Plasmid

<400> 121

accatgaaaa	catttaacat	ttctcaacaa	gatctagaat	tagtagaagt	agcgacagag	60
aagattacaa	tgctttatga	ggataataaa	catcatgtgg	gagcggcaat	tcgtacgaaa	120
acaggagaaa	tcatttcggc	agtacatatt	gaagcgtata	taggacgagt	aactgtttgt	180
gcagaagcca	ttgcgattgg	tagtgcagtt	tcgaatggac	aaaaggattt	tgacacgatt	240
gtagctgtta	gacacctta	ttctgacgaa	gtagatagaa	gtattcgagt	ggtaagtcct	300
tgtgggtatgt	gtagggagtt	gatttcagac	tatgcaccag	attgttttgt	gttaatagaa	360
atgaatggca	agttagtcaa	aactacgatt	gaagaactca	ttccactcaa	atataccga	420
aaattaaaga	tttaccatac	caagcttggc	tgctgcctga	ggctggacga	cctcgcggag	480
ttctaccggc	agtgcacaa	cgtcggcatc	caggaaacca	gcagcggcta	tcgcgcgcatc	540
catgcccccg	aactgcagga	gtggggaggc	acgatggccg	ctttggtcgg	gatctttgtg	600
aaggaacctt	acttctgtgg	tgtgacataa	ttggacaaa	tacctacaga	gatttaaacg	660
tctaaggtaa	atataaaatt	tttaagtgt	taatgtgtta	aactactgat	tctaattgtt	720
tgtgtatttt	agattccaac	ctatggaact	gatgaatggg	agcagtgggtg	gaatgccttt	780
aatgaggaaa	acctgttttg	ctcagaagaa	atgccatcta	gtgatgatga	ggctactgct	840
gactctcaac	attctactcc	tccaaaaaag	aagagaaagg	tagaagaccc	caaggacttt	900
ccttcagaat	tgctaagttt	tttgagtcac	gctgtgttta	gtaatagaac	tcttgcttgc	960
tttgctattt	acaccacaaa	ggaaaaagct	gcactgctat	acaagaaaa	tatggaaaaa	1020
tattctgtaa	cctttataag	taggcataac	agttataatc	ataacatact	gttttttctt	1080
actccacaca	ggcatagagt	gtctgctatt	aataactatg	ctcaaaaatt	gtgtaccttt	1140
agctttttta	tttgtaaagg	ggttaataag	gaatatttga	tgtatagtgc	cttgactaga	1200
gatcataatc	agccatacca	catttgtaga	ggttttactt	gctttaaaaa	acctcccaca	1260
cctccccctg	aacctgaaac	ataaaaatga	tgcaattgtt	gttggttaact	tgttttattgc	1320
agcttataat	ggttacaaat	aaagcaatag	catcacaaat	ttcacaaata	aagatccaga	1380
tttcgctcaa	gttagtataa	aaaagcaggc	ttcaatcctg	cagagaagct	tggcgccagc	1440
cggcttcaat	tgacggggcc	ccactagtga	gtcgtattac	gtagcttggc	gtaatcatgg	1500
tcatagctgt	ttcctgtgtg	aaattgttat	ccgctcacaa	ttccacacaa	catacgagcc	1560
ggaagcataa	agtgtaaagc	ctggggtgcc	taatgagtga	gctaactcac	attacatgtg	1620
agcaaaaggc	cagcaaaagg	ccaggaaaccg	taaaaaggcc	gcgttgctgg	cgttttttcca	1680
taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtcaga	ggtggcgaaa	1740
cccgacagga	ctataaagat	accaggcggt	tccccctgga	agctccctcg	tgcgctctcc	1800
tgttccgacc	ctgccgctta	ccggataacct	gtccgccttt	ctcccttcgg	gaagcgtggc	1860
gctttctcat	agctcacgct	gtaggatatc	cagttcggtg	taggtcggtc	gctccaagct	1920
gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	1980
tcttgagtc	aacccggtaa	gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	2040
gattagcaga	gcgaggtatg	taggcggtgc	tacagagttc	ttgaagtgg	ggcctaacta	2100
cggtacacat	agaagaacag	tatttgggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	2160
aaaaagagtt	ggtagctctt	gatccggcaa	acaaaaccacc	gctggtagcg	gtgggtttttt	2220
tgtttgcaag	cagcagatta	cgcgcagaaa	aaaaggatct	caagaagatc	ctttgatctt	2280
ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	taagggattt	tggcatagag	2340
attatcaaaa	aggatcttca	cctagatcct	tttacgcgcc	ctgtagcggc	gcattaagcg	2400
cggcggtgtg	ggtggttacg	cgcagcgtga	ccgctaacat	tgccagcgcc	ctagcgcccg	2460
ctcctttcgc	tttcttccct	tcctttctcg	ccacgttcgc	tttcccgcgc	aagctctaaa	2520
tcgggggctc	ccttttaggt	tccgatttag	tgctttacgg	cacctcgacc	ccaaaaaact	2580
tgatttgggt	gatggttcac	gtagtggggc	atcgccctga	tagacgggtt	ttcgcccttt	2640
gacgttggag	tccacgttct	ttaatagtgg	actcttggtc	caaactggaa	caacactcaa	2700
ccctatctcg	ggctattctt	ttgatttata	agggattttg	ccgatttcgg	cctattgggt	2760
aaaaaatgag	ctgatttaac	aaaaatttaa	cgcgaatttt	aacaaaatat	taacgtttac	2820
aatttaataa	tttgcttata	caatcttcct	gtttttgggg	cttttctgat	tatcaaccgg	2880
ggtaaatcaa	tctaaagtat	atatgagtaa	acttggtctg	acagttacca	atgcttaatc	2940
agtgaggcac	ctatctcagc	gatctgtcta	tttctgtcat	ccatagttgc	ctgactcccc	3000
gtcgtgtaga	taactacgat	acgggagggc	ttaccatctg	gccccagtcg	tgcaatgata	3060
ccgcgagacc	cacgctcacc	ggctccagat	ttatcagcaa	taaaccagcc	agccggaagg	3120
gccgagcgca	gaagtgttcc	tgcaacttta	tccgcctoca	tccagtctat	taattgttgc	3180
cgggaagcta	gagtaagtag	ttcgccagtt	aatagtttgc	gcaacgttgt	tgccatttgt	3240

-91-

```

acaggcatcg tgggtgtcacg ctcgtcgttt ggtatggctt cattcagctc cggttcccaa 3300
cgatcaaggc gagttacatg atcccccatg ttgtgcaaaa aagcgggttag ctcccttcgg 3360
cctccgatcg ttgtcagaag taagttggcc gcagtgttat cactcatggg tatggcagca 3420
ctgcataatt ctcttactgt catgccatcc gtaagatgct tttctgtgac tggtgagtac 3480
tcaaccaagt cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgtca 3540
acacgggata ataccgcgcc acatagcaga actttaaaag tgctcatcat tggagaacgt 3600
tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaacct 3660
actcgtgcac ccaactgatc ttcagcatct tttactttca ccagcgtttc tgggtgagca 3720
aaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa atgttgaaata 3780
ctcatactct tcctttttca atattattga agcattttatc aggggttattg tctcatgagc 3840
ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc 3900
cgaaaagtgc cacctgacgt agttaacaaa ccaagcccg cgaagcggg ctttattacc 3960
aagcgaagcg ccattcgcca ttcaggctgc gcaactgttg ggaagggcga tcggtgcggg 4020
cctcttcgct attacgccag ctggcgaaag ggggatgtgc tgcaaggcga ttaagttggg 4080
taacgccagg gttttcccag tcacgacggt gtaaaacgac ggccagtcg taatacgact 4140
cacttaaggc cttgactaga ggtcgacgc atgcctgtac atccggagac gcgtcacggc 4200
cgaagctagc gaattcgtgg atc 4223

```

<210> 122

<211> 2686

<212> DNA

<213> Artificial Sequence

<220>

<223> pUC18 Plasmid

<400> 122

```

tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cctaactatg cggcatcaga gcagattgta ctgagagtgc 180
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc 240
attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcgggccc tcttcgctat 300
tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt aagttgggta acgccagggt 360
tttcccagtc acgacgttgt aaaacgacgg ccagtgccaa gcttgcatgc ctgcaggctc 420
actctagagg atccccggg accgagctcg aattcgtaat catggtcata gctgtttcc 480
gtgtgaaatt gttatccgct cacaattcca cactcaaac gagccggaag cataaagtgt 540
aaagcctggg gtgcctaatt agtgagctaa ctacattaa ttgcgttgcg ctactgccc 600
gctttccagt cgggaaacct gtctgcccag ctgcattaat gaatcggcca acgcgcgggg 660
agaggcgggt tcgctattgg gcgctcttcc gcttccctgc tactgactc gctgcgctcg 720
gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtaatacg gttatccaca 780
gaatcagggg ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa ggccagggaac 840
cgtaaaaagg ccgcgttgct ggctgttttc cataggctcc gccccctga cgagcatcac 900
aaaaatcgac gctcaagtca gaggtggcga aaccgcgacg gactataaag ataccaggcg 960
tttccccctt gaagctccct cgtgcgctct cctgttccga ccctgcgct taccggatac 1020
ctgtccgctt ttctcccttc gggaagcgtg cgcgtttctc atagctcacg ctgtaggtag 1080
ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc cccggttcag 1140
cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaaccgggt aagacacgac 1200
ttatcgccac tggcagcagc cactggtaac aggtattgca gagcgaggta tgtaggcgg 1260
gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttgg 1320
atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc 1380
aaacaaacca ccgctggtag cgggtggttt tttgtttgca agcagcagat tacgcgcaga 1440
aaaaagggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc tcagtggaac 1500
gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt cacctagatc 1560
cttttaaat aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggctc 1620
gacagttacc aatgcttaat cagtgaagca cctatctcag cgatctgtct atttcgttca 1680
tccatagttg cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct 1740
ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca 1800
ataaacccag cagccggaag ggcgagcgc agaagtggc ctgcaacttt atccgcctcc 1860
atccagctca ttaattgttg ccgggaagct gttcgccagt gttcgccagt taatagttt 1920
cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggtagggc 1980
tcattcagct ccggttccca acgatcaagg cgagttacat gatcccccat gttgtgcaaa 2040
aaagcggtta gctccttcgg tcctccgatc gttgtcagaa gtaagttggc cgcagtgtta 2100
tactcatgga ttatggcagc actgcataat tctcttactg tcatctagag ctgaagatgc 2160
ttttctgtga ctggtgagta ctcaaccaag taattgtgat aatagtgat gcggcgaccg 2220
agttgctctt gcccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa 2280
gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg 2340
agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc 2400

```

-92-

accagcgttt	ctgggtgagc	aaaaacagga	aggcaaaatg	ccgcaaaaaa	gggaataagg	2460
gcgacacgga	aatggtgaat	actcatactc	ttcctttttc	aatattattg	aagcatttat	2520
cagggttatt	gtctcatgag	cggatacata	tttgaatgta	tttagaaaaa	taaacaaata	2580
gggggtccgc	gcacatttcc	ccgaaaagtg	ccacctgacg	tctaagaaac	cattatttatc	2640
atgacattaa	cctataaaaa	taggcgtatc	acgaggccct	ttcgtc		2686

<210> 123

<211> 8521

<212> DNA

<213> Artificial Sequence

<220>

<223> pCXeGFPattB(6xHS4)2 Plasmid

<400> 123

tacggggcgg	gggatccact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccataat	ggagttccgc	gttacataac	ttacggtaaa	tggcccgccct	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtac	240
atcaagtgt	tcatatgcc	agtacgcccc	ctattgacgt	caatgacggg	aaatggcccc	300
cctggcatta	tgcccagtac	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattagtcat	cgctattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctcccc	420
atctcccccc	cctccccacc	cccaattttg	tatttattta	ttttttaatt	attttgtgca	480
gcgatggggg	cggggggggg	ggggggcgcg	gccaggcggg	gcggggcggg	gcgagggggc	540
ggggcgggcg	aggcgggag	gtgcgggcg	agccaatcag	agcgggcgcg	tccgaaagtt	600
tccttttatg	gcgagggcg	ggcgggcgcg	gccctataaa	aagcgaagcg	cgcgggcggg	660
gggagtcgct	gcgttgccct	cgccccgtgc	cccgctccgc	gccgcctcgc	gcccggcgcc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtagcgggg	cgggacggcc	cttctcctcc	780
gggctgtaat	tagcgcttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgcgtagaaag	840
ccttaaaggg	ctccgggagg	gcccttttgt	cgggggggag	cggtccgggg	ggtgctgctg	900
tgtgtgtgtg	ctgtggggag	gccgcgtgcg	gcccgcgctg	cccgccggct	gtgagcgctg	960
cgggcgcgcg	gcggggcttt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggccgggggc	1020
ggtgccccgc	ggtgcggggg	ggctgcgagg	ggaacaaagg	ctgctgctcg	ggtgtgtgctg	1080
tgggggggtg	agcagggggg	gtggggcgcg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cgggccggct	tcgggtgctg	ggctccgtgc	ggggcggtgg	1200
gcggggctcg	ccgtgcccgg	cggggggtgg	cggcaggtgg	gggtgcccgg	cggggcgggg	1260
ccgcctcggg	ccgggggagg	ctcgggggag	gggcgcggcg	gccccggagc	gcccggcggt	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttatggta	atcgctgcgag	agggcgcgagg	1380
gacttccttt	gtcccaaata	tggcggagcc	gaaatctggg	aggcgccgc	gcacccccct	1440
tagcgggcgc	gggcgaagcg	gtgcggcgcc	ggcaggaagg	aaatgggcgg	ggagggccct	1500
cgtgcgtcgc	cgcgccgcgc	tcccctctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acggctgcct	tcggggggga	cggggcaggg	cggggttcgg	cttctggcgt	gtgaccggcg	1620
gctctagagc	ctctgctaac	catgttcatt	ccttcttctt	tttctacacg	ctcctgggca	1680
acgtgctggt	tgttgtgtgt	tctcatcatt	ttggcaaaga	attcgccacc	atggtgagca	1740
agggcgagga	ggttttcacc	gggggtggtg	ccatcctggt	cgagctggac	ggcgacgtaa	1800
acggccacaa	gttcagcgtg	tcgggcgagg	gcgagggcga	tgccacctac	ggcaagctga	1860
ccctgaagtt	catctgcacc	accggcaagc	tgcccgtgcc	ctggcccacc	ctcgtgacca	1920
ccctgacct	cggcgtgcag	tgcttcagcc	gctaccccca	ccacatgaag	cagcacgact	1980
tcttcaagtc	cgccatgccc	gaaggctacg	tcaggagcgc	caccatcttc	ttcaaggacg	2040
acggcaacta	caagaccgcg	gccgaggtga	agttcgaggg	cgacaacctg	gtgaaccgca	2100
tcgagctgaa	gggcatcgac	ttcaaggagg	acggcaacat	cctggggcac	aagctggagt	2160
acaactacaa	cagccacaa	gtctatatca	tggccgacaa	gcagaagaac	ggcatcaagg	2220
tgaacttcaa	gatccgccac	aacatcgagg	acggcagcgt	gcagctcgcc	gaccactacc	2280
agcagaacac	ccccatcggc	gacggccccc	tgctgctgcc	cgacaaccac	tacctgagca	2340
cccagtcgcg	cctgagcaaa	gaccccaacg	agaagcgga	tcacatgggt	ctgctggagt	2400
tcgtgaccgc	cgccgggata	actctcggca	tgagcagcgt	gtacaagtaa	gaattcactc	2460
ctcaggtgca	ggctgcctat	cagaaggtgg	tggtggtggt	ggccaatgcc	ctggctcaca	2520
aataccactg	agatcttttt	cctctcgcca	aaaattatgg	ggacatcatg	aagccccttg	2580
agcatctgac	ttctggctaa	taaaggaaat	ttattttcat	tgcaatagtg	tgttgggaat	2640
ttttgtgtct	ctcactcgga	aggacatatg	ggaggggcaa	tcatttataa	catcagaatg	2700
agtatttgg	ttagagtttg	gcaacatatg	ccatatgctg	gctgccatga	acaaaggtgg	2760
ctataaagag	gtcatcagta	tatgaaacag	ccccctgctg	tccattccct	attccataga	2820
aaagccttga	cttgagggtta	gatttttttt	atattttgg	ttgtgttatt	tttttcttta	2880
acatccctaa	aattttccct	acatgtttta	ctagccagat	tttttctcct	ctcctgacta	2940
ctccagtcga	tagctgtccc	tcttctctta	tgaagatccc	tcgacctgca	gcccagcttt	3000
ggcgtaatca	tgggtcatagc	tgtttccctg	gtgaaattgt	tatccgctca	caattccaca	3060
caacatacga	gccggaagca	taaagtgtaa	agcctggggg	gcctaattgag	tgagctaact	3120

cacattaatt	gcgttgccgt	cactgcccgc	tttccagtcg	ggaaacctgt	cgtgccagcg	3180
gatccgcac	tcaattagtc	agcaaccata	gtcccgcgcc	taactccgcc	catcccgcgc	3240
ctaactccgc	ccagttccgc	ccattctccg	ccccatgggt	gactaatttt	ttttatttat	3300
gcagaggccg	aggccgcctc	ggcctctgag	ctattccaga	agtagtgagg	aggctttttt	3360
ggaggctagt	ggatcccccg	ccccgtatcc	cccaggtgtc	tgcaggctca	aagagcagcg	3420
agaagcggtc	agaggaaaag	gatcccggtc	caccttcccc	gtgcccgggc	tgtcccgcga	3480
cgctgcccgc	tcggggatgc	ggggggagcg	ccggaccgga	gcccagcccc	ggcggtctcg	3540
ctgctgcccc	ctagcggggg	agggaacgtaa	ttacatccct	gggggctttg	gggggggggt	3600
gtccccgtga	gcggatccgc	ggccccgtat	cccccagggt	tctgcaggct	caaagagcag	3660
cgagaagcgt	tcagaggaaa	gcgatcccggt	gccaccttcc	ccgtgcccgg	gctgtccccg	3720
cacgctgccc	gctcggggat	gcggggggag	cgccggaccg	gagcggagcc	ccgggaggct	3780
cgctgctgcc	ccctagcggg	ggaggacgtg	aattacatcc	ctggggggct	tggggggggg	3840
ctgtccccgt	gagcggatcc	gcggccccgt	atcccccagg	tgtctgcagg	ctcaaagagc	3900
agcgagaagc	gttcagagga	aagcgatccc	gtgccacctt	ccccgtgccc	gggctgtccc	3960
cgcacgctgc	cggctcgggg	atgccccggg	agcgcgggac	cggagcggag	ccccggggcg	4020
ctcgtgctgc	ccccctagcg	ggggagggac	gtaattacat	ccctgggggc	tttggggggg	4080
ggctgtcccc	ccgcggcccc	ccgcggcccc	gtatccccca	gggtgtctga	ggctcaaaga	4140
gcagcgagaa	gcgttcagag	gaaagcgatc	ccgtgccacc	ttccccgtgc	ccgggctgtc	4200
cccgcacgct	gccggctcgg	ggatgcgggg	ggagcgcggg	accggagcgg	agccccgggc	4260
ggctcgtctg	tgccccctag	cgggggaggg	acgtaattac	atccctgggg	gctttggggg	4320
ggggctgtcc	ccgtgagcgg	atccgcggcc	ccgtatcccc	cagggtgtctg	caggctcaaa	4380
gagcagcgag	aagcgttcag	aggaaaagcga	tcccggtcca	ccctccccgt	gccccgggctg	4440
tccccgcacg	ctgccggctc	ggggatgcgg	ggggagcggc	ggaccggagc	ggagccccgg	4500
gcggctcgtc	gctgccccct	agcgggggag	ggacgtaatt	acatccctgg	gggctttggg	4560
gggggggctgt	ccccgtgagc	ggatccgcgg	ccccgtatcc	cccagggtgtc	tgcaggctca	4620
aagagcagcg	agaagcgttc	agaggaaaag	gatcccggtc	caccttcccc	gtgcccgggc	4680
tgtccccgca	cgtgccccgc	tcggggatgc	ggggggagcg	ccggaccgga	gcccagcccc	4740
gggcggctcg	ctgctgcccc	ctagcggggg	agggaacgtaa	ttacatccct	gggggctttg	4800
gggggggggt	gtccccgtga	gcggatccgc	ggggctgcag	gaattcgatt	gaagcctgtt	4860
ttttttatact	aacttgagcg	aaatcaagct	cctaggcttt	tgcaaaaagc	taacttggtt	4920
attgcagctt	ataatggtta	caaataaaag	aatagcatca	caaatttcac	aaataaaagc	4980
ttttttttcac	tgcatcttag	ttgtggtttg	tccaaaactca	tcaatgtatc	ttatcatgtc	5040
tggatccgct	gcattaatga	atcggccaac	gcgcggggag	aggcggtttg	cgtattgggc	5100
gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggg	cggttcggctg	cggcgagcgg	5160
tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	aacgcaggaa	5220
agaacatgtg	agcaaaaagg	cagcaaaaagg	ccaggaaaccg	taaaaaggcc	gcgttgctgg	5280
cgttttttcca	taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtcaga	5340
ggtggcgaaa	cccgcacagga	ctataaaagat	accaggcggt	tccccctgga	agctccctcg	5400
tgcgctctcc	tgttccgacc	ctgcccgtta	ccggatacct	gtccgccttt	ctcccttcgg	5460
gaagcgtggc	gctttctcaa	tgctcacgct	gtaggatctc	cagttcggtg	taggtcggtc	5520
gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	5580
gtaactatcg	tcttgagttc	aaccgggtaa	gacacgactt	atcgccactg	gcagcagcca	5640
ctggtaacag	gatttagcaga	gcgaggtatg	taggcgggtg	tacagagttc	ttgaagtggg	5700
ggcctaacta	cggctacact	agaaggacag	tatttggtat	ctgcgctctg	ctgaagccag	5760
ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtacgc	5820
gtgggttttt	tgtttgcaag	cagcagatta	cgcgcagaaa	aaaaggatct	caagaagatc	5880
ctttgatctt	ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	taagggtatt	5940
tggatcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	aatgaagttt	6000
ttaaatcaat	ctaaagtata	tatgagtaaa	cttgggtctga	cagttaccaa	tgcttaatca	6060
gtgaggcacc	tatctcagcg	atctgtctat	ttcggttcac	catagttgcc	tgactccccg	6120
tcgtgtagat	aactacgata	cgggaggggt	taccatctgg	ccccagtgct	gcaatgatac	6180
cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	gccggaaggg	6240
ccgagcgcag	aagtggctct	gcaactttat	ccgcctccat	ccagtctatt	aattggttgc	6300
gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	6360
caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	atcagctccc	ggttcccaac	6420
gatcaaggcg	agttacatga	tcccccatgt	tgtgcaaaaa	agcgggttagc	tccttcggtc	6480
ctccgatcgt	tgtcagaagt	aagttggccg	cagtgttatc	actcatgggt	atggcagcac	6540
tgcataattc	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	6600
caaccagtcg	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgtaaa	6660
tacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacggt	6720
cttcggggcg	aaaactctca	aggatcttac	cgctgttgag	atccagttcg	atgtaaccca	6780
ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	6840
aaacagggaag	gcaaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	tggtgaatac	6900
tcatactctt	cccttttcaa	tattattgaa	gattttatca	gggttattgt	ctcatgagcg	6960
gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	7020
gaaaagtgcc	acctgggtcga	cggtatcgat	aagcttgata	togaattcct	gcagccccgc	7080
ggatccgctc	acgggggacag	ccccccccc	aagccccag	ggatgtaatt	acgtccctcc	7140

-94-

```

ccccgctaggg ggcagcagcg agccgccccg ggtccgctc cggtcggcg cccccccgc 7200
atccccgagc cggcagcgtg cggggacagc ccgggcacgg ggaaggtggc acgggatcgc 7260
tttcctctga acgcttctcg ctgctctttg agcctgcaga cacctggggg atacggggcc 7320
gcggatccgc tcacggggac agcccccccc caaagcccc agggatgtaa ttacgtccct 7380
cccccgctag ggggcagcag cgagccgccc ggggctccgc tccggtcggg cgtcccccc 7440
gcatccccga gccggcagcg tgcggggaca gccgggcac ggggaaggtg gcacgggatc 7500
gctttcctct gaacgcttct cgctgctctt tgagcctgca gacacctggg ggatacgggg 7560
ccggcgatcc gctcacgggg acagcccccc cccaaagccc ccagggatgt aattacgtcc 7620
ctcccccgct agggggcagc agcagccgc ccggggctcc gctccgggtc ggcgctcccc 7680
ccgcatcccc gagccggcag cgtgcgggga cagcccgggc acgggggaagg tggcacggga 7740
tcgctttcct ctgaacgctt ctgctgctc tttagacctg cagacacctg ggggatacgg 7800
ggccgcggat ccgctcacgg gcacagcccc ccccaaagc cccagggat gtaattacgt 7860
ccctcccccg ctagggggca gcagcgagcc gcccggggct ccgctccggc ccggcgctcc 7920
ccccgcatcc ccgagccggc agcgtgcggg gacagcccg gcacggggaa ggtggcacgg 7980
gatcgctttc ctctgaacgc ttctcgctgc tctttgagcc tgcagacacc tgggggatac 8040
ggggccgcgg atccgctcac ggggacagcc ccccccaaa gccccaggg atgtaattac 8100
gtccctcccc ccgtaggggg cagcagcgag ccgcccgggg ctccgctccg gtcggcgct 8160
ccccccgcat ccccgagccg gcagcgtgcg gggacagccc gggcacgggg aaggtggcac 8220
gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca cctgggggat 8280
acggggccgc ggatccgctc acggggacag cccccccca aagccccag ggatgtaatt 8340
acgtccctcc ccgctaggg ggacgagcg agccgcccgg ggctccgctc cggtcggcg 8400
ctccccccgc atccccgagc cggcagcgtg cggggcacgg ccggggcacgg ggaaggtggc 8460
acgggatcgc tttcctctga acgcttctcg ctgctctttg agcctgcaga cacctggggg 8520
a 8521

```

<210> 124

<211> 8851

<212> DNA

<213> Artificial Sequence

<220>

<223> p18EPOcDNA Plasmid

<400> 124

```

cagttgccgg ccgggtcgcg cagggcgaac tcccgcccc acggctgctc gccgatctcg 60
gtcatggccg gcccgaggc gtcccggaag ttctgtgaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccg caccacacc caggccaggg tggtgtccgg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtcccga ccacaccggc gaagtcgtcc 240
tccacgaagt cccgggagaa cccgagccgg tcgggtccaga actcgaccgc tccggcgacg 300
tcgcgcgagg tgagcacgg aacggcactg ctcaacttgg ccattggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatata gaattcctgc 420
agcccccgcg atccgctcac ggggacagcc ccccccaaa gccccaggg atgtaattac 480
gtccctcccc cgctaggggg cagcagcgag ccgcccgggg ctccgctccg gtcggcgct 540
ccccccgcat ccccgagccg gcagcgtgcg gggacagccc gggcacgggg aaggtggcac 600
gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca cctgggggat 660
acggggccgc ggatccgctc acggggacag cccccccca aagccccag ggatgtaatt 720
acgtccctcc ccgctaggg ggacgagcg agccgcccgg ggctccgctc cggtcggcg 780
ctccccccgc atccccgagc cggcagcgtg cggggacagc ccggggcacgg ggaaggtggc 840
acgggatcgc tttcctctga acgcttctcg ctgctctttg agcctgcaga cacctggggg 900
atacggggcc gcggatccgc tcacggggac agcccccccc caaagcccc agggatgtaa 960
ttacgtccct ccccgctag ggggcagcag cgagccgccc ggggctccgc tccggtcggg 1020
cgctcccccc gcatccccga gccggcagcg tgcggggaca gcccgggcac ggggaaggtg 1080
gcacgggatc gctttcctct gctcacgggg acagcccccc cccaaagccc ccagggatgt 1200
ggatacgggg ccgcgatcc gctcacgggg agcagccgc ccggggctcc gctccgggtc 1260
aattacgtcc ctccccgct agggggcagc cgtgcgggga cagcccgggc acgggggaagg 1320
ggcgctcccc ccgcatccc ctgaacgctt ctgctgctc tttgagcctg cagacacctg 1380
tggaacggga ggcccgggat ccgctcacgg ggacagcccc ccccaaagc cccagggat 1440
gttaattacgt cctcccccg cttagggggc cagcagagcc gcccggggct ccgctccgg 1500
ccggcgctcc ccccgcatcc ccgagccggc agcgtgcggg gacagcccgg gcacggggaa 1560
ggtggcacgg gatcgcttct ctctgaacgc ttctcgctgc tctttgagcc tgcagacacc 1620
tgggggatac gggggccggc atccgctcac ggggacagcc ccccccaaa gccccaggg 1680
atgtaattac gtccctcccc cgctaggggg cagcagcgag ccgcccgggg ctccgctccg 1740
gtccggcgct ccccgcatcc ccccgagccg gggacagccc gggacagccc cctgcagaca 1800
aaggtggcac gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca 1860
cctgggggat acggggcggg ggatccacta gttattaata gtaatcaatt acgggggtcat 1920
tagttcatag cccatatatg gagttccgcg ttacataact tacggtaaat ggcccgccctg 1980

```

gctgaccgccc	caacgaccccc	cgccccattga	cgtcaataaat	gacgtatggtt	cccatagtaa	2040
cgccaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgcccact	2100
tggcagtaga	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggccccc	ctggcattat	gcccagtaca	tgaccttatg	ggacttttct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tgggtcgagg	tgagccccac	gttctgcttc	2280
actctcccca	tctccccccc	ctccccaccc	ccaattttgt	atltatlttat	tttttaatta	2340
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cgggcggggg	2400
cgagggggcg	ggcgggggcg	ggcgggagag	tgcgggcgga	gccaatcaga	gcggcgcgct	2460
ccgaaagttt	ccttttatgg	cgaggcgggc	gcgggcgggc	ccctataaaa	agcgaagcgc	2520
gcgggcgggc	ggagtcgctg	cgttgccctc	gccccgtgcc	ccgctccgcg	ccgcctcgcg	2580
ccgcccgcgc	cggctctgac	tgaccgcgtt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctcctccg	ggctgttaatt	agcgcttgg	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
gcgtgaaagc	cttaaagggc	tccgggaggg	ccctttgtgc	gggggggagc	ggctcggggg	2760
gtgctgtgct	gtgtgtgtgc	gtggggagcg	ccgcgtgctg	ccgcgctgct	ccggcggtcg	2820
tgagcgctgc	ggggcgggcg	cggggctttg	tgcgctccgc	gtgtgcgcga	ggggagcgcg	2880
gcccggggcg	gtgccccgcg	gtgcgggggg	gctgcgaggg	gaacaaagcg	tcgctgcggg	2940
gtgtgtgctg	gggggggtga	gcagggggtg	tgggcgcggc	ggtcgggctg	taaccccccc	3000
ctgcaccccc	ctccccaggt	tgtgagcac	ggcccggtt	cgggtgcggg	gctccgtgcg	3060
gggcgtggcg	cggggctcgc	cgtgcggggc	gggggggtgg	ggcaggtggg	ggtgccgggc	3120
ggggcggggc	cgcctcgggc	cggggagggc	tccggggagg	ggcgcgggcg	ccccggagcg	3180
ccggcggtcg	tcgagcgcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
ggggcgaggg	acttcccttg	tcccaaatct	ggcggagccg	aaatctggga	ggcgccgcgc	3300
cacccccctc	agcgggcgcg	ggcgaagcgg	tgcgcgcccg	gcaggaagga	aatgggcggg	3360
gagggccttc	gtgctgcgc	gcgcgcgcct	ccccttctcc	atctccagcc	tcggggctgc	3420
cgcaggggga	cggctgcctt	cgggggggac	ggggcagggc	gggggttcggc	ttctggcgctg	3480
tgaccggcgg	ctctagaatg	gggggtgcacg	aatgtcctgc	ctggctgtgg	cttctcctgt	3540
ccctgctgtc	gtccctctg	ggcctcccag	tcctgggcgc	cccaccacgc	ctcatctgtg	3600
acagccgagt	cctggagagg	tacctcttgg	aggccaagga	ggccgagaat	atcacgacgg	3660
gctgtgctga	acactgcagc	ttgaatgaga	atatcactgt	cccagacacc	aaagttaatt	3720
tctatgcctg	gaagaggatg	gaggtcgggc	agcaggccgt	agaagtctgg	cagggcctgg	3780
ccctgctgtc	ggaagctgtc	ctgcggggcc	aggccctggt	ggtaactct	tcccagcgtg	3840
gggagccccc	gcagctgcac	gtggataaag	ccgtcagtgg	ccttcgcagc	ctcaccactc	3900
tgcttcgggc	tctgggagcc	cagaagggaag	ccatctcccc	tccagatgcg	gcctcagctg	3960
ctccactccg	aacaatcact	gctgacactt	tccgcaaaact	cttcagagtc	tactccaatt	4020
tccctcgggg	aaagctgaag	ctgtacacag	gggagggcctg	caggacaggg	gacagtgac	4080
gtacaagtaa	gaattcactc	ctcaggtgca	ggctgcctat	cagaagggtg	tggctggtgt	4140
ggccaatgcc	ctggctcaca	aataccactg	agatcttttt	cctctgccca	aaaattatgg	4200
ggacatcatg	aagccccctg	agcatctgac	ttctggctaa	taaaggaaat	ttatlttcat	4260
tgcaatagtg	tggttgaatt	ttttgtgtct	ctcactcgga	aggacatatg	ggagggcaaa	4320
tcatttaaaa	catcagaatg	agtaatttgg	ttagagtgtg	gcaacatatg	ccatatgctg	4380
gctgccatga	acaaagggtg	ctataaagag	gtcactcagta	tatgaaacag	ccccctgctg	4440
tccattccct	attccataga	aaagccttga	cttgagggtta	gatttttttt	atattttgtt	4500
ttgtgttatt	tttttcttta	acatccctaa	aattttccct	acatgtttta	ctagccagat	4560
ttttcctcct	ctcctgacta	ctcccagtca	tagctgtccc	tcttctctta	tgaagatccc	4620
tcgacctgca	gccccagctt	gcatgcctgc	aggtgcactc	tagtggatcc	cccgcccgct	4680
atcccccagg	tgtctgcagg	ctcaaagagc	agcgagaagc	gttcagagga	aagcgatccc	4740
gtgccaccct	ccccgtgccc	gggctgtccc	cgcacgctgc	cggctcgggg	atgcgggggg	4800
agcgccggac	cggagcggag	ccccggggcg	ctcgctgctg	ccccctagcg	ggggagggag	4860
gtaattacat	ccttgggggg	tttggggggg	ggctgtcccc	gtgagcggat	ccgcggcccc	4920
gtatccccca	gggtgtctga	ggctcaaaag	gcagcgagaa	gcgttcagag	gaaagcgatc	4980
ccgtgccacc	ttccccgtgc	ccgggctgtc	cccgcacgct	gcgggctcgg	ggatgcgggg	5040
ggagcgccgg	acgggagcgg	agccccgggc	ggctcgctgc	tgccccctag	cgggggaggg	5100
acgtaattac	atccctgggg	gctttggggg	ggggctgtcc	ccgtgagcgg	atccgcggcc	5160
ccgtatcccc	caggtgtctg	caggtcctaa	gagcagcgag	aagcgttcag	aggaaaagcga	5220
tcccgtgcca	ccttccccgt	gcccgggctg	tccccgcacg	ctgcgggctc	ggggatgcgg	5280
ggggagcgcc	ggaccggagc	ggagcccccg	gcggctcgct	gctgccccct	agcgggggag	5340
ggacgtaatt	acatccctgg	gggctttggg	ggggggctgt	ccccgtgagc	ggatccgcgg	5400
ccccgtatcc	cccaggtgtc	tgcaggctca	aagagcagcg	agaagcgttc	agaggaaagc	5460
gatcccgctg	caccttcccc	gtgcccgggc	cgtgccccga	cgtgccccgc	tcggggatgc	5520
ggggggagcg	ccggaccgga	gcggagcccc	gggcgggctc	ctgctgcccc	ctagcggggg	5580
agggacgtaa	ttacatccct	gggggctttg	ggggggggct	gtccccgtga	gcggatccgc	5640
ggccccgtat	cccccaggtg	tctgcaggct	caaagagcag	cgagaagcgt	tcagaggaaa	5700
gcgatcccg	gccacccttc	ccgtgccccg	gctgtccccg	cacgtgccc	gctcggggat	5760
gcggggggag	cgcgggacgg	cgggcgggct	cgggcgggct	cgtgctgccc	ccctagcggg	5820
ggagggagct	aattacatcc	ctgggggctt	tggggggggg	ctgtccccgt	gagcggatcc	5880
gcgggccccg	atcccccagg	tgtctgcagg	ctcaaagagc	agcgagaagc	gttcagagga	5940
aagcgatccc	gtgccacctt	ccccgtgccc	gggctgtccc	cgcacgctgc	cggctcgggg	6000

-96-

```

atgcgggggg agcgccggag cggagcggag ccccgggggg ctcgctgctg cccctagcgg 6060
ggggaggggac gtaattacat ccctgggggg tttggggggg ggctgtcccc gtgagcggat 6120
ccgcgggggct gcaggaattc gtaatcatgg tcatagctgt ttctgtgtgt aaattgttat 6180
ccgctcacaa ttccacacaa catacagagc ggaagcataa agtgtaaagc ctgggggtgcc 6240
taatgagtga gctaactcac attaatgtcg ttgcgctcac tgcccgcctt ccagtcggga 6300
aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgctg 6360
attgggcgct cttccgcttc ctgcgtcact gactcgctgc gtcgggtcgt tcgggtgcgg 6420
cgagcgggat cagctcactc aaaggcggta atacggttat ccacagaatc aggggataaac 6480
gcaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg 6540
ttgctggcgt ttttccatag gctccgcccc cctgacgagc atcacaaaaa tcgacgctca 6600
agtccagagt ggcgaaaccc gacaggacta taaagatacc aggcgttttc ccctggaagc 6660
tcctcgtgc gctctcctgt tccgacctgc ccgcttacgg gatacctgtc cgctttctc 6720
ccttcgggaa gcgtggcgct ttctcatagc tcacgctgta ggtatctcag ttcggtgtag 6780
gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg ttcagcccgga ccgctgcgcc 6840
ttatccggta actatcgtct tgagtccaac ccggtaagac acgacttatc gccactggca 6900
gcagccactg gtaacaggat tagcagagcg aggtatgtag gcggtgctac agagttcttg 6960
aagtgggtggc ctaactacgg ctacactaga aggacagtat ttggtatctg cgctctgctg 7020
aagccagtta ccttcggaaa aagagttggg agctcttgat ccggcaaaaa aaccaccgct 7080
ggtagcggtg gtttttttgt ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa 7140
gaagatcctt tgatcttttc tacggggtct gacgctcagt ggaacgaaaa ctcacgttaa 7200
gggatttttg tcatgagatt atcaaaaagg atcttcacct agatcctttt aaattaaaaa 7260
tgaagtttta aatcaatcta aagtatatat gagtaaacct ggtctgacag ttaccaatgc 7320
ttaatcagtg aggcacctat ctacgcatc tgtctatttc gttcatccat agttgcctga 7380
ctccccgtcg ttagataaac tacgatacgg gagggccttac catctggccc cagtgctgca 7440
atgataccgc gagaccacag ctacccggct ccagatttat cagcaataaa ccagccagcc 7500
ggaagggccg agcgacagaag tggctctgca actttatccg cctccatcca gtctattaat 7560
tgttcgcggg aagctagagt aagtagttcg ccagttaata gtttgcgcaa cgttgttgcc 7620
attgctacag gcatcgtggg gtcacgctcg tcgtttggta tggcttcatt cagctccggg 7680
tcccaacgat caaggcagat tacatgatcc cccatgttgt gcaaaaaaagc ggttagctcc 7740
ttcggtcctc cgatcgttgt cagaagtaag ttggccgcag tgttatcact catggttatg 7800
gcagcactgc ataattctct tactgtcatg ccatccgtaa gatgcttttc tgtactggg 7860
gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc gaccgagttg ctcttgccc 7920
gcgtaataac gggataatac cgcgccacat agcagaactt taaaagtgc catcattgga 7980
aaacgttctt cggggcgaaa actctcaagg atcttacccg tgttgagatc cagttcgatg 8040
taaccactc gtgcacccaa ctgatcttca gcatctttta ctttcaccag cgtttctggg 8100
tgagcaaaaa caggaaggca aatgcccga aaaaagggaa taagggcgac acggaaatgt 8160
tgaatactca tactcttct tttcaatat tattgaagca tttatcaggg ttattgtctc 8220
atgagcggat acatatttga atgtatttag aaaaataaac aaataggggt tccgcgcaca 8280
tttccccgaa aagtgccacc tgacgtagtt aacaaaaaaa agcccgcga agcgggctt 8340
attaccaagc gaagcgccat tcgccattca ggctgcgcaa ctggtgggaa ggcgatccg 8400
tgccggcttc ttgcgtatta cgccagctgg cgaaaggggg atgtgctgca aggcgatcaa 8460
gttgggtaac gccagggttt tcccagtcac gacgttgtaa aacgacggcc agtccgtaat 8520
acgactcact taaggccttg actagagggt cgacgggtata cagacatgat aagatacatt 8580
gatgagtttg gacaaaccac aactagaatg cagtgaaaaa aatgctttat ttgtgaaatt 8640
tgtgttgcta ttgctttatt tgtaaccatt ataactgca ataaacaagt tgggtgggg 8700
gaagaactcc agcatgagat ccccgcgctg gaggatcatc cagccggcgt cccggaaaac 8760
gattccgaag cccaacctt catagaaggc ggcggtggaa tcgaaatctc gtagcacgtg 8820
tcagtcctgc tcctcggcca cgaagtgcac g

```

<210> 125

<211> 10474

<212> DNA

<213> Artificial Sequence

<220>

<223> p18genEPO Plasmid

<400> 125

```

cagttgccgg ccgggtcgcg cagggcgaa ccccccccc acggctgctc gccgatctcg 60
gtcatggcgg gcccggaggc gtcccgaag ttcgtggaca cgacctcga ccactcggcg 120
tacagctcgt ccaggccgcg caccacacc caggccaggg tgttgtccgg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtccccga ccacaccggc gaagtcttcc 240
tccacgaagt cccgggagaa cccgagccgg tcggtccaga actcgaccgc tccggcgacg 300
tcgcgcggcg tgagcaccgg aacggcactg tgcgaacttg ccatggatcc agatttcgt 360
caagttagta taaaaagca ggcttcaatc ctgcagagaa gcttgatatc gaattcctgc 420
agcccccgcg atccgctcac ggggacagcc cccccc aaa gccccaggg atgtaattac 480
gtccctcccc cgctaggggg cagcagcgag ccgccccggg ctcgctccg gtccggcgct 540

```

ccccccgcat	ccccgagccg	gcagcgtgcg	gggacagccc	gggcacgggg	aaggtggcac	600
gggagcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	cctgggggat	660
acggggccgc	ggatccgctc	acggggacag	cccccccca	aagccccag	ggatgtaatt	720
acgtccctcc	cccgctaggg	ggcagcagcg	agccgcccgg	ggctccgctc	cggtccggcg	780
ctccccccgc	atccccgagc	cggcagcgtg	cggggacagc	ccgggcacgg	ggaaggtggc	840
acgggatcgc	tttctcttga	acgttctctg	ctgctctttg	agcctgcaga	cacctggggg	900
atacggggcc	gcggatccgc	tcacggggac	agcccccccc	caaagcccc	agggatgtaa	960
ttacgtccct	ctcccgcctag	ggggcagcag	cagaccgccc	ggggctccgc	tccggtccgg	1020
cgctcccccc	gcatccccga	gccggcagcg	tgccggggaca	gcccgggcac	ggggaagggtg	1080
gcacgggatc	gctttctctt	gaacgcttct	cgctgctctt	tgagcctgca	gacacctggg	1140
ggatacgggg	ccgcggatcc	gctcacgggg	acagcccccc	cccaaagccc	ccagggatgt	1200
aattacgtcc	ctcccgcgtc	agggggcagg	agcgagccgc	ccggggctcc	gctccggccc	1260
ggcgctcccc	ccgcatcccc	gagccggcag	cgtgcgggga	cagcccgggc	acgggggaagg	1320
tggcacggga	tcgctttcct	ctgaacgctt	ctcgctgctc	tttgagcctg	cagacacctg	1380
ggggatacgg	ggcgcgggat	ccgctcacgg	ggacagcccc	ccccaaaagc	ccccagggat	1440
gtaattacgt	ccctcccccg	ctagggggga	gcagcgagcc	gcccggggct	ccgctccggt	1500
ccggcgctcc	ccccgcattc	ccgagccggc	agcgtgcggg	gacagcccg	gcacggggaa	1560
ggtggcacgg	gatcgctttc	ctctgaacgc	ttctcgctgc	tctttgagcc	tgacagacacc	1620
tgggggatac	ggggccgcgg	atccgctcac	ggggacagcc	ccccccaaa	gcccccagg	1680
atgtaattac	gtccctcccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	1740
gtccggcgct	ccccccgcat	ccccgagcgg	gcagcgtgcg	gggacagccc	gggcacgggg	1800
aagtgggcac	gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttattaata	gtaatcaatt	acgggggtcat	1920
tagttcatag	cccatatatg	gagttccgcg	ttacataact	tacggtaaat	ggcccgcctg	1980
gctgaccgcc	caacgacccc	cgccccattga	cgtcaataat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgccact	2100
tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagtaga	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tgggtcgagg	tgagccccac	gttctgcttc	2280
actctcccca	tctccccccc	ctccccaccc	ccaattttgt	atttatttat	tttttaatta	2340
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cggggcgggg	2400
cgagggggcg	ggcggggcga	ggcgggagag	tgccggcgga	gccaatcaga	gcggcgcgct	2460
ccgaaagttt	ccttttatgg	cgaggcgggc	gcggcgggcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtgcgtg	cggttgcctt	gccccgtgcc	ccgctccgcg	ccgcctcgcg	2580
ccgcccgcgc	cggtctgac	tgaccgcgtt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctcctccg	ggctgtaat	agcgttggt	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
gcgtgaaagc	cttaaagggc	tccgggaggg	ccctttgtgc	gggggggagc	ggctcggggg	2760
gtgcgtgcgt	gtgtgtgtgc	gtggggagcg	ccgcgtgcgg	cccgcgctgc	ccggcgggctg	2820
tgagcgctgc	gggcgcggcg	cggggctttg	tgcgctccgc	gtgtgcgcga	ggggagcgcg	2880
gcccggggcg	gtgcccgcg	gtgcgggggg	gctgcgaggg	gaacaaaggc	tgcggtcggg	2940
gtgtgtgcgt	gggggggtga	gcaggggggt	ggggcgggc	ggtcgggctg	taaccccccc	3000
ctgcaccccc	ctccccgagt	tgctgagcac	ggcccggtt	cgggtgcggg	gctccgtgcg	3060
gggcgtggcg	cggggctcgc	cgtgcggggc	ggggggtggc	ggcaggtggg	ggtgcccggg	3120
ggggcggggg	cgcctcgggc	cggggagggc	tccggggagg	ggcgcgggcg	ccccggagcg	3180
ccggcgcgctg	tcgagcgcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
gggcgcaggg	acttcctttg	tcccaaatct	ggcggagccg	aaatctggga	ggcgccgccc	3300
cacccccctct	agcgggcgcg	ggcgaagcgg	tgccggcgcc	gcaggaagga	aatgggcggg	3360
gagggccttc	gtgcgtcgcc	gcgcgcgcgt	ccccttctcc	atctccagcc	tcgggggtgc	3420
cgcaggggga	cggtgcctt	cgggggggac	ggggcagggc	ggggttcggc	ttctggcggtg	3480
tgaccggcg	ctctagatgc	atgctcgagc	ggccgccagt	gtgatggata	tctgcagaat	3540
tcgccctttc	tagaatgggg	gtgcacgggtg	agtactcgcg	ggctgggcgc	tcccgcccg	3600
ccgggtccct	gtttgagcgg	ggatttagcg	ccccggctat	tgccagagg	gtgggtgggt	3660
tcaaggaccg	gcgacttgtc	aaggaccccg	gaagggggag	gggggtgggg	tgccctccacg	3720
tgccagcggg	gacttggggg	agtccttggg	gatggcaaaa	acctgacctg	tgaaggggac	3780
acagtttggg	ggttgagggg	aagaaggttt	gggggttctg	ctgtgccagt	ggagaggaag	3840
ctgataagct	gataacctgg	gcgctggagc	caccacttat	ctgccagagg	ggaagcctct	3900
gtcacaccag	gattgaagtt	tggccggaga	agtggatgct	ggtagctggg	ggtgggggtg	3960
gcacacggca	gcaggattga	atgaaggcca	gggagggcag	acctgagtgc	ttgcatgggt	4020
ggggacagga	aggacgagct	ggggcagaga	cgtggggatg	aaggaaagctg	tccttcacaa	4080
gccacccttc	tccctccccg	cctgactctc	agcctggcta	tctgttctag	aatgtcctgc	4140
ctggctgtgg	cttctcctgt	cctgctgtgc	gctccctctg	ggcctcccag	tcctggggcg	4200
cccaccacgc	ctcatctgtg	acagccgagt	cctggagagg	tacctcttgg	aggccaagga	4260
ggccgagaat	atcacggtga	gaccccttcc	cctgcacatt	ccacagaact	caggtccagg	4320
gcttcaggga	actcctccca	gatccaggaa	ccaggcaact	gggttggggg	ggagtgggga	4380
agctagacac	tgccccctta	cataagaata	agtctgggtg	ccccaaacca	tacctggaaa	4440
ctaggcaagg	agcaaagcca	gcagatccta	cggcctgtgg	gccagggcca	gagccttcag	4500
ggacccttga	ctccccgggc	tgtgtgcatt	tcagacgggc	tgtgctgaac	actgcagctt	4560

-98-

gaatgagaat	atcactgtcc	cagacaccaa	agttaatttc	tatgcctgga	agaggatgga	4620
ggtgagttcc	tttttttttt	tttttccctt	cttttggaga	atctcatttg	cgagcctgat	4680
tttggatgaa	agggagaatg	atcgagggaa	aggtaaaatg	gagcagcaga	gatgagggctg	4740
cctgggcgca	gaggctcacg	tctataatcc	caggctgaga	tggccgagat	gggagaattg	4800
cttgagccct	ggagtttcag	accaacctag	gcagcatagt	gagatccccc	atctctacaa	4860
acatttaaaa	aaattagtca	ggtgaagtgg	tgcattggtg	tagtcccaga	tatttggaag	4920
gctgagggcg	gagatcgct	tgagcccagg	aatttgaggc	tgagtgagc	tgtgatcaca	4980
ccactgcact	ccagcctcag	tgacagagt	aggccctgtc	tcaaaaaaga	aaagaaaaaa	5040
gaaaaataat	gagggctgta	tggaatacat	tcattattca	ttcactcact	cactcactca	5100
ttcatttcatt	cattcattca	acaagtctta	ttgcatacct	tctgtttgct	cagcttgggtg	5160
cttggggctg	ctgaggggca	ggaggggag	ggtgacatgg	gtcagctgac	tcccagagtc	5220
cactccctgt	aggtcgggca	gcaggcctga	gaagctctgg	agggcctggc	cctgctgtcg	5280
gaagctgtcc	tgcgggggca	ggccctgttg	gtcaactctt	cccagccgtg	ggagcccctg	5340
cagctgcatg	tggataaagc	cgtcagtggt	cttcgcagcc	tcaccactct	gcttcgggct	5400
ctgggagccc	aggtgagtag	gagcggacac	ttctgcttgc	cctttctgta	agaaggggag	5460
aagggctctt	ctaaggagta	caggaactgt	ccgtattcct	tccctttctg	tggcactgca	5520
gcgacccatt	gttttctcct	tggcagaagg	aagccatctc	ccctccagat	gcggcctcag	5580
ctgctccact	ccgaacaatc	actgctgaca	ctttccgcaa	actcttcoga	gtctactcca	5640
atcttccctc	gggaaagctg	aagctgtaca	caggggaggg	ctgcaggaca	ggggacagat	5700
gacgtacaag	taagaattca	ctcctcaggt	gcaggtctgc	tatcagaagg	tgggtggctgg	5760
tgtggccaat	gccctggctc	acaaaataca	ctgagatctt	tttccctctg	ccaaaaatta	5820
tggggacatc	atgaagcccc	ttgagcatct	gacttctggc	taataaagga	aattttatttt	5880
cattgcaata	gtgtgttgga	attttttgtg	tctctcactc	ggaaggacat	atgggagggc	5940
aaatcattta	aaacatcaga	atgagtatct	ggttttagagt	ttggcaacat	atgccatatg	6000
ctggctgcca	tgaacaaagg	tggtataaaa	gaggtcatca	gtatatgaaa	cagccccctg	6060
ctgtccattc	cttattccat	agaaaagcct	tgacttgagg	ttagattttt	tttatatttt	6120
gttttgtgtt	atttttttct	ttaacatccc	taaaattttc	cttacatgtt	ttactagcca	6180
gatttttctc	cctctcctga	ctactcccag	tcatactgtg	ccctcttctc	ttatgaagat	6240
ccctcgacct	gcagcccaag	cttgcatgoc	tgcaggtcga	ctctagtggg	tcccccgccc	6300
cgtatccccc	aggtgtctgc	aggtctaaag	agcagcgaga	agcgttcaga	ggaaagcgat	6360
cccgtgccac	cttcccctg	cccgggctgt	ccccgcacgc	tgccggctcg	gggatgcggg	6420
gggagcgccg	gaccggagcg	gagccccggg	cggctcgctg	ctgcccccta	gcgggggagg	6480
gacgtaatca	catccctggg	ggctttgggg	gggggctgtc	cccgtgagcg	gatccgcggc	6540
cccgtatccc	ccaggtgtct	gcaggtctca	agagcagcga	gaagcgttca	gaggaaagcg	6600
atcccctg	accttcccc	tgcccgggct	gtccccgcac	gctgcccggc	cggggatgoc	6660
gggggagcgc	cggaccggag	cggagccccg	ggcggtcgc	tgctgcccc	tagcggggga	6720
gggacgtaat	tacatccctg	ggggctttgg	gggggggctg	tccccgtgag	cggatccgcg	6780
gccccgtatc	ccccaggtgt	ctgcaggctc	aaagagcagc	gagaagcgtt	cagaggaaag	6840
cgatcccggt	ccaccttccc	cgtgccccgg	ctgtccccgc	acgctgccgg	ctcggggatg	6900
cggggggagc	gcccggagcc	agcggagccc	cgggcggtc	gctgctgccc	cctagcggg	6960
gagggagcgt	attacatccc	tgggggcttt	gggggggggc	tgtccccgtg	agcggatccg	7020
cggccccgta	tccccaggt	gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	7080
agcgatcccc	tgccaccttc	cccgtgccc	ggctgtcccc	gcacgctgcc	ggctcggggg	7140
tgcgggggga	gcgccggacc	ggagcggagc	ccggggcggc	tcgctgctgc	cccctagcgc	7200
gggagggagc	taattacatc	cctgggggct	ttgggggggg	gctgtcccc	tgagcggatc	7260
cgcggccccg	tatccccag	gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	7320
aaagcgatcc	cgtgccacct	tcccgtgccc	cgggctgtcc	ccgcacgctg	ccggctcggg	7380
gatgcggggg	gagcgccgga	ccggagcggg	gccccggggc	gctcgctgct	gccccctagc	7440
gggggagggg	cgtaatatac	tccctggggg	ctttgggggg	gggctgtccc	cgtgagcggg	7500
tccgcggccc	cgtatcccc	aggtgtctgc	aggctcaaa	agcagcgaga	agcgttcaga	7560
ggaaagcgat	cccgtgccac	cttcccctg	cccgggctgt	ccccgcacgc	tgccggctcg	7620
gggatgcggg	gggagcgccg	gaccggagcg	gagccccggg	cggctcgctg	ctgcccccta	7680
gcgggggagg	gacgtaatca	catccctggg	ggctttgggg	gggggctgtc	cccgtgagcg	7740
gatccgcggg	gctgcaggaa	ttcgtaatac	tggctcatagc	tgtttctgtg	gtgaaatttg	7800
tatccgctca	caattccaca	caacatacga	gccggaagca	taaaagttaa	agcctggggg	7860
gcctaataag	tgagctaact	cacattaatt	gcgttgctgc	cactgcccgc	tttccagtcg	7920
ggaaacctgt	cgtgccagct	gcattaatga	atcgcccaac	gcgcggggag	agggcggttg	7980
cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgctcctggg	cgttcggctg	8040
cggcgagcgg	ctcaaaggcg	ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	8100
aacgcaggaa	agaacatgtg	agcaaaaggc	cagcaaaagg	ccaggaaacc	taaaaaggcc	8160
gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacia	aaatcgacgc	8220
tcaagtacga	ggtggcgaaa	cccgcagagg	ctataaagat	accaggcggt	tccccctgga	8280
agctccctcg	tgcgctctcc	tgctccgacc	ctgcgcctta	ccggatacct	gtccgccttt	8340
ctcccttcgg	gtaagctggc	gctttctcat	agctcacgct	gtaggtatct	cagttcgggtg	8400
taggtcgctc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	8460
gccttatccg	gtaactatcg	tcttgagtc	aaccgcgtaa	gacacgactt	atcgccactg	8520
gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	taggcgggtg	tacagagttc	8580

-99-

ttgaagtgggt	ggcctaacta	cggctacact	agaaggacag	tatttggtat	ctgcgctctg	8640
ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	8700
gctggtagcg	gtggtttttt	tgtttgcaag	cagcagatta	cgcgagaaa	aaaaggatct	8760
caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaacga	aaactcacgt	8820
taagggattt	tgggtcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	8880
aaatgaagtt	ttaaatcaat	ctaaagtata	tatgagtaaa	cttgggtctga	cagttacca	8940
tgcttaatca	gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcac	catagttgcc	9000
tgactccccg	tcgtgtagat	aactacgata	cgggagggct	taccatctgg	ccccagtgct	9060
gcaatgatac	cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	9120
gccggaaggg	ccgagcgcag	aagtggctct	gcaactttat	ccgcctccat	ccagtctatt	9180
aattggtgcc	gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	9240
gccattgcta	caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	9300
ggttcccaac	gatcaaggcg	agttacatga	tccccatgt	tgtgcaaaaa	agcgggttagc	9360
tccttcggtc	ctccgatcgt	tgtcagaagt	aagttggccg	cagtgttatc	actcatgggt	9420
atggcagcac	tgcataattc	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	9480
gggtgagta	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgtcttgc	9540
cggcgctcaa	tacgggataa	taccgcgcca	catagcagaa	ctttaaagt	gctcatcatt	9600
ggaaaacgtt	cttcggggcg	aaaactctca	aggatcttac	cgctgttgag	atccagttcg	9660
atgtaaccca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	9720
gggtgagcaa	aaacaggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	9780
tgttgaatac	tcatactctt	cctttttcaa	tattattgaa	gcatttatca	gggttattgt	9840
ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	ggttccgcgc	9900
acattttccc	gaaaagtgcc	acctgacgta	gttaacaaaa	aaaagcccg	cgaagcgggc	9960
tttattacca	agcgaagcgc	cattcgccat	tcaggctgcg	caactgttgg	gaagggcgat	10020
cgggtgcggg	ctcttcgcta	ttacgccagc	tggcgaaagg	gggatgtgct	gcaaggcgat	10080
taagttgggt	ttttcccagt	cacgacgttg	taaaacgacg	gccagtcagt	10140	
aatacgactc	acttaaggcc	ttgactagag	ggtcgacggg	atacagacat	gataagatac	10200
attgatgagt	ttggacaaac	cacaactaga	atgcagtgaa	aaaaatgctt	tatttgtgaa	10260
atltgtgatg	ctattgcttt	atltgttaacc	attataagct	gcaataaaca	agttgggggtg	10320
ggcgaagaac	tccagcatga	gatccccgcg	ctggaggatc	atccagccgg	cgtcccggaa	10380
aacgattccg	aagcccaacc	tttcatagaa	ggcgccgggtg	gaatcgaaat	ctcgtagcac	10440
gtgtcagtc	tgctcctcgg	ccacgaagtg	cacg			10474

<210> 126

<211> 6119

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeoeGFP Plasmid

<400> 126

cagttgccgg	ccgggtcgcg	cagggcggaac	tcccgccccc	acggctgctc	gccgatctcg	60
gtcatggcgc	ccccggaggc	gtcccgaag	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggccgcg	cacccacacc	caggccaggg	tgttgtccgg	caccacctgg	180
tcctggaccg	cgctgatgaa	cagggtcacg	tcgtcccggg	ccacaccggc	gaagtgcgtc	240
tcacagaaat	ccggggagaa	cccagaccgg	tcggtccaga	actcgaccgc	tcggggcgacg	300
tcgctgcgcg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgggctg	caggtcgagg	420
gatcttcata	agagaagagg	gacagctatg	actgggagta	gtcaggagag	gaggaaaaat	480
ctggctagta	aaacatgtaa	ggaaaatttt	agggatgtta	aagaaaaaaa	taacacaaaa	540
caaaatataa	aaaaaatcta	acctcaagtc	aaggcttttc	tatggaataa	ggaatggaca	600
gcagggggct	gtttcatata	ctgatgacct	ctttatagcc	acctttgttc	atggcagcca	660
gcataatggca	tatgtttgcca	aactotaaac	caaatactca	ttctgatgtt	ttaaatgatt	720
tgccctccca	tatgtccttc	cagagtgaag	acacaaaaaa	ttccaacaca	ctattgcaat	780
gaaaataaat	ttcctttatt	agccagaagt	cagatgctca	aggggcttca	tgatgtcccc	840
ataatttttg	gcagagggaa	aaagatctca	gtggatattg	tgagccaggg	cattggccac	900
accagccacc	accttctgat	aggcagcctg	cacctgagga	gtgaattctt	acttgtacag	960
ctcgtccatg	ccgagagtga	tcccggcggc	gctcacgaac	tccagcagga	ccatgtgatc	1020
gcgcttctcg	ttggggctct	tgctcagggc	ggactgggtg	ctcaggtagt	ggttgtcggg	1080
cagcagcacg	gggcccgtcg	cgatgggggt	gttctgctgg	tagtggctcg	cgagctgcac	1140
gctgccgtcc	tcgatgttgt	ggcggtatct	gaagttcacc	ttgatgccgt	tcttctgctt	1200
gtcggccatg	atagagcgtg	tgtggctgtt	gtagttgtac	tccagcttgt	gccccaggat	1260
gttgccgtcc	tccttgaagt	cgatgccctt	cagctcgatg	cggttcacca	gggtgtcgcc	1320
ctcgaacttc	acctcggcgc	gggtcttgta	gttgccgtcg	tccttgaaga	agatgggtcg	1380
ctcctggacg	tagccttcgg	gcattggcga	cttgaagaag	tcgtgctgct	tcatgtgggt	1440
ggggtagcgg	ctgaagcact	gcacgccgta	ggtcaggggtg	gtcacgaggg	tgggccaggg	1500

-100-

cacggggcagc	ttgccgggtg	tgcagatgaa	cttcaggggtc	agcttgccgt	aggtggcatc	1560
gcccccgccc	tgcggcgaca	cgctgaactt	gtggcggttt	acgtcgccgt	ccagctcgac	1620
caggatgggg	accaccccg	tgaacagctc	ctcgcccttg	ctcaccatgg	tggcgaattc	1680
tttgccaaaa	tgatgagaca	gcacaacaac	cagcacgttg	cccaggagct	gtaggaaaaa	1740
gaagaaggca	tgaacatggt	tagcagaggc	tctagagccg	ccggtcacac	gccagaagcc	1800
gaaccccgcc	ctgccccgtc	cccccggaag	gcagccgtcc	ccctgcggca	gccccgaggc	1860
tggagatgga	gaaggggacg	gcggcgccgc	gacgcacgaa	ggccctcccc	gccccatttc	1920
ttcctgccc	cgccgcaccg	cttcgcccgc	gcccgttaga	gggggtgcgg	cggcgcctcc	1980
cagatttcgg	ctccgccaga	tttgggacaa	aggaagtccc	tgcgccctct	cgcacgatta	2040
ccataaaagg	caatggctgc	ggctcgccgc	gctcgcacag	ccgcccggcg	tccggggccg	2100
ccgcgcccc	cccccgagcc	ctccccggcc	cgaggcgcc	ccgccccgcc	cggcaccccc	2160
acctgcccgc	accccccgcc	cggcacggcg	agcccccgcc	cacgccccgc	acggagcccc	2220
gcacccgaag	ccggggccgtg	ctcagcaact	cggggagggg	ggtgcagggg	ggggttacag	2280
cccgaccgcc	gcgcccacac	ccctgtctca	ccccccacg	cacacacccc	gcacgcagcc	2340
tttgttcccc	tgcagcccc	cccgccacgc	ggggcaccgc	ccccggccgc	gctccccctc	2400
cgcacacgcg	gagcgacaaa	agcccccgcc	cgcgcccgca	gcgctcacag	ccgcggcgca	2460
gcgcgggccg	cacgcggcgc	tccccacgca	cacacacacg	cacgcacccc	ccgagccgct	2520
cccccccgca	caaagggccc	tcccgaggcc	ctttaaggct	ttcacgcagc	cacagaaaag	2580
aaacgagccg	tcatataaac	aagcgcta	tacagcccgg	aggagaagg	ccgtcccgc	2640
cgctcacctt	tgggagtaac	gcggtcagtc	agagccgggg	cgggcggcgc	gaggcgccgc	2700
ggagcggggg	acggggcgaa	ggcaacgcag	cgactcccgc	ccgcccgcgc	cttcgctttt	2760
tatagggccg	ccgcccgcgc	cgcctcgcca	taaaaggaaa	ctttcggagc	gcgcccgtct	2820
gattggctgc	cgccgcacct	ctccgcctcg	ccccgccccg	cccctcgccc	cgccccgccc	2880
cgctggcg	cgcccccccc	cccccccgcc	cccatcgctg	cacaaaataa	ttaaaaaata	2940
aataaaatac	aaattggggg	tggggagggg	ggggagatgg	ggagagtga	gcagaaacgtg	3000
gggtcacctt	cgacccatgg	taatagcgat	gactaatacg	tagatgtact	gccaagttag	3060
aaagtcccat	aaggtcatgt	actgggcata	atgccaggcg	ggccatttac	cgctattgac	3120
gtcaatagg	ggcgtaactt	gcataatgata	cacttgatgt	actgccaagt	gggcagttta	3180
ccgtaaatag	tccaccattt	gacgtcaatg	gaaagtccct	attggcggtt	ctatgggaac	3240
atacgtcatt	attgacgtca	atgggcgggg	gtcgttgggc	ggtcagccag	gcgggccatt	3300
taccgtaagt	tatgtaacgc	ggaactccat	atatgggcta	tgaactaatg	accccgtaac	3360
tgattactat	taataactag	aggatccccg	ggtaccgagc	tcgaattcgt	aatcatggtc	3420
atagctgttt	cctgtgtgaa	attgttatcc	gctcaccaat	ccacacaaca	tacgagccgg	3480
aagcataaag	tgtaaagcct	ggggtgccta	atgagtgcgc	taactcacat	taattgcgtt	3540
gcgctcaact	cccgctttcc	agtccggaaa	ctgtcgctgc	cagctgcatt	aatgtaacgg	3600
ccaacgcgcg	gggagaggcg	gtttgcgtat	tgggcgctct	tccgcttcc	cgctcactga	3660
ctcgctgcgc	tccggtcgttc	ggctgcggcg	agcggtatca	gctcactcaa	aggcggtaat	3720
acggttatcc	acagaatcag	gggataacgc	aggaaagaac	atgtgagcaa	aaggccagca	3780
aaaggccagg	aaccgtaaaa	aggccgcgtt	gctggcggtt	ttccataggc	tccgcccccc	3840
tgacgagcat	cacaaaaatc	gacgctcaag	tcagagggtg	cgaaacccga	caggactata	3900
aagataccag	gcgtttcccc	ctggaagctc	cctcgtgcgc	tctcctgttc	cgaccctgcc	3960
gcttaccgga	tacctgtccg	cctttctccc	ttcgggaagc	gtggcgcttt	ctcatagctc	4020
acgctgtagg	tatctcagtt	cgggtgtagt	cgttcgctcc	aagctgggct	gtgtgcacga	4080
accccccggt	cagcccgacc	gctgcgcctt	atccggtaac	tatcgtcttg	agtcacaacc	4140
ggtaagacac	gacttatcgc	cactggcagc	agccactgg	aacaggatta	gcagagcgag	4200
gtatgtaggc	ggtgctacag	agtctctgaa	gtggtggcct	aactacggct	acactagaag	4260
gacagtattt	ggtatctgcg	ctctgctgaa	gccagttacc	ttcggaaaaa	gagttggtag	4320
ctcttgatcc	ggcaaaaaag	ccaccgctgg	tagcgggtgg	ttttttgttt	gcaagcagca	4380
gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	atcttttcta	cggggtctga	4440
cgctcagtg	aacgaaaact	cacgttaagg	gattttgggtc	atgagattat	caaaaaggat	4500
cttcacctag	atccttttaa	attaaaaatg	aagttttaaa	tcaatctaaa	gtatatatga	4560
gtaaaacttg	tctgacagtt	accaatgctt	aatcagtgag	gcacctatct	cagcgatctg	4620
tctatctcgt	tcatccatag	ttgctgact	ccccgctgtg	tagataacta	cgatacggga	4680
gggcttacca	tctggcccca	gtgctgcaat	gataccgcga	gaccacgcct	caccggctcc	4740
agatttatca	gcaataaacc	agccagccgg	aaggcccgag	cgcagaagtg	gtcctgcaac	4800
tttatccgcc	tccatccagt	ctattaattg	ttgcccggaa	gctagagtaa	gtagttcgcc	4860
agttaatagt	ttgcgcaacg	ttgttgccat	gtctacaggc	atcgtgggtg	cacgctcgct	4920
gtttgggtatg	gctccggttc	ttagctcctt	ccaacgatca	aggcgagtta	catgacccc	4980
catgttgtgc	aaaaaagcgg	ttagctcctt	cggctcctcg	atcgttgtca	gaagtaagtt	5040
ggccgcagtg	ttatcactca	tgggtatggc	agcactgcat	aattctctta	ctgtcatgcc	5100
atccgtaaga	tgcctttctg	tgactggtga	gtactcaacc	aagtcattct	gagaatatgtg	5160
tatcgggcga	ccgagttgct	cttgcccgcg	gtcaataccg	gataataccg	cgccacatag	5220
cagaacttta	aaagtgtctca	tcattggaaa	acgttcttcg	gggcgaaaaa	tctcaaggat	5280
cttaccgctg	ttgagatcca	gttcgatgta	acccactcgt	gcacccaact	gatcttcagc	5340
atcttttact	ttcaccagcg	tttctgggtg	agcaaaaaa	ggaaggcaaa	atgccgcaaa	5400
aaagggaata	agggcgacac	ggaaatgttg	aatactcata	ctcttccttt	ttcaatatata	5460
ttgaagcatt	tatcagggtt	attgtctcat	gagcggatac	atatttgaat	gtatttagaa	5520

-101-

```

aaataaacaac atagggggttc cgcgcacatt tccccgaaaa gtgccacctg acgtagttaa 5580
caaaaaaaag cccgcggaag cgggctttat taccaagcga agcgccattc gccattcagg 5640
ctgcgcaact gttgggaagg gcgatcgggtg cgggcctctt cgctattacg ccagctggcg 5700
aaaggggggat gtgctgcaag gcgattaagt tgggtaacgc cagggttttc ccagtcacga 5760
cgttgtaaaa cgacggccag tccgtaatac gactcactta aggccttgac tagagggtcg 5820
acggtataca gacatgataa gatacattga tgagtttgga caaaccacaa ctagaatgca 5880
gtgaaaaaaa tgctttatct gtgaaatttg tgatgctatt gctttatctg taaccattat 5940
aagctgcaat aaacaagttg ggggtgggcga agaactccag catgagatcc ccgcgctgga 6000
ggatcatcca gccggcgctcc cggaaaaacga ttccgaagcc caacctttca tagaaggcgg 6060
cgggtggaatc gaaatctcgt agcacgtgtc agtctgtctc ctcggccacg aagtgcacg 6119

```

<210> 127

<211> 5855

<212> DNA

<213> Artificial Sequence

<220>

<223> pCXLamInt Plasmid (Wildtype Integrase)

<400> 127

```

gtcgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60
gcccatatat ggagttccgc gttacataac ttacggtaaa tggccgcgct ggctgaccgc 120
ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180
ggacttttcca ttgacgtcaa tgggtggact atttacggta aactgccacac ttggcagtag 240
atcaagtgtga tcatatgcca agtacgcccc ctattgacgt caatgacggg aaatggcccc 300
cctggcatta tgcccagtag atgaccttat gggactttcc tacttggcag tacatctacg 360
tattagtcac cgctattacc atgggtcgag gtgagcccca cgttctgctt cactctcccc 420
atctcccccc cctccccacc cccaattttg tatttattta ttttttaatt attttgtgca 480
gcatggtggg cggggggggg gggggcgcg cccaggcggg gcggggcggg gcgagggggc 540
gggcggggcg aggcgggag gtgcccggc agccaatcag agcggcgcgc tccgaaagt 600
tccttttatg gcgaggcggc ggcggcggc gccctataaa aagcgaagcg gcgcggcggc 660
gggagtcgct gcgttgccct cgcgccgtgc cccgctccgc gccgcctcgc gcgcgccgcc 720
ccggctctga ctgaccgcgt tactcccaca ggtgagcggg cgggacggcc cttctcctcc 780
gggtgtgaat tagcgcttgg ttaaatgacg gctcgtttct tttctgtggc tgcgtgaaag 840
ccttaaaagg ctccgggagg gccctttgtg cgggggggag cggctcgggg ggtgcgtgcg 900
tgtgtgtgtg cgtggggagc gccgcgtgcg gcccgcgctg cccggcggtg gtgagcgctg 960
cgggcgcggc gcggggcttt gtgcgctccg cgtgtgcgcg aggggagcgc ggccgggggc 1020
gggtccccgc ggtgccccgg ggctgcgagg ggaacaaagg ctgctgctgc ggtgtgtgcg 1080
tgggggggtg agcagggggt gtgggcgcgg cggtcgggct gtaaccccc cctgcacccc 1140
cctccccgag ttgctgagca cggcccggtc tcgggtgcgg ggctccgtgc ggggcgtggc 1200
gcggggctcg ccgtgcgggg cggggggtgg cggcaggtgg ggggtgcggg cggggcgggg 1260
ccgcctcggg ccggggaggg ctccgggggag gggcgcggcg gccccggagc gccggcggtc 1320
gtcgaggcgc ggcgagccgc agccattgac ttttatggta atcgtgcgag agggcgcgag 1380
gacttccttt gtcccaaatc tggcggagcc gaaatctggg aggcgcgcgc gcaccccc 1440
tagcgggcgc gggcgaagcg gtgcggcgcc ggcaggaagg aaatggcgcg ggagggcctt 1500
cgtgcgtcgc cgcgcgcgcg tccccttctc catctccagc ctcggggctg ccgcaggggg 1560
acggctgcct tcggggggga cggggcaggg cgggggttcg cttctggcgt gtgaccggcg 1620
gctctagagg ctctgctaac catgttcatt ccttcttctt ttctctacag ctccctggga 1680
acgtgctggt tgttgtgctg tctcatcatt ttggcaaaga attcatggga agaaggcgaa 1740
gtcatgagcg ccgggattta cccctaacc tttatataag aaacaatgga tattactgct 1800
acagggaccc aaggacgggt aaagagtttg gattaggcag agacaggcga atcgcaatca 1860
ctgaagctat acaggccaac attgagttat ttccaggaca caaacacaag cctctgacag 1920
cgagaatcaa cagtataat tccgttacgt tacattcatg gcttgatcgc tacgaaaaaa 1980
tcctggccag cagaggaaat aagcagaaga cactcataaa ttacatgagc aaaattaaag 2040
caataaggag gggctctgcct gatgctccac ttgaagacat caccacaaaa gaaattgcgg 2100
caatgctcaa tggatacata gacgagggca aggcggcgct ctgaaggcca atcagatcaa 2160
cactgagcga tgcattccga gaggcaatag cagaggttaa acttacggct aaccatgtcg 2220
ctgccactcg cgcagcaaaa tcagaggtaa ggaggtcagg acttacggct gacgaatacc 2280
tgaaaattta tcaagcagca gaatcatcac catgttggct cagacttgca atggaactgg 2340
ctgttggttac cgggcaacga gttggtgatt tatgcgaaat gaagtggctc gatatcgtag 2400
atggatatct ttatgtcgag caaagcaaaa caggcgtaaa aattgccatc ccaacagcat 2460
tgcatattga tgctctcgga atatcaatga aggaacact tgataaatgc aaagagattc 2520
ttggcggaga aaccataatt gcatctactc ctgcgcaacc gctttcatcc ggcacagtat 2580
caaggtatct tatgcgcgca cgaaaagcat caggctcttc cttcgaaggg gatccgccta 2640
cctttcacga gttgcgcagt ttgtctgcaa gactctatga gaagcagata agcgataagt 2700
ttgctcaaca tcttctcggg cataagtcgg acaccatggc atcacagtat cgtgatgaca 2760
gaggcagggg gtgggacaaa attgaaatca aataagaatt cactcctcag gtgcaggctg 2820

```

-102-

```

cctatcagaa ggtggtggct ggtgtggcca atgccctggc tcacaaatac cactgagatc 2880
tttttccctc tgccaaaaaat tatggggaca tcatgaagcc ccttgagcat ctgacttctg 2940
gctaataaag gaaattttatt ttcatgtcaa tagtgtgttg gaattttttg tgtctctcac 3000
tcggaaggac atatgggagg gcaaatcatt taaaacatca gaatgagtat ttggtttaga 3060
gtttggcaac atatgccata tgctggctgc catgaacaaa ggtggctata aagaggtcat 3120
cagtatatga aacagccccc tgctgtccat tccttattcc atagaaaagc cttgacttga 3180
ggttagattt tttttatatt ttgtttgtg ttattttttt ctttaacatc cctaaaattt 3240
tccttacatg ttttactagc cagatttttc ctctctctct gactactccc agtcatagct 3300
gtccctcttc tcttatgaag atccctcgac ctgcagccca agcttggcgt aatcatggtc 3360
atagctgttt cctgtgtgaa attgttatcc gctcacaatt ccacacaaca tacgagccgg 3420
aagcataaag tgtaaagcct ggggtgccta atagtgagc taactcacat taattgcgtt 3480
gcgctcactg cccgcttttc agtcgggaaa cctgtcgtgc cagcggatcc gcatctcaat 3540
tagtcagcaa ccatagtccc gccctaaact ccgcccctac cgcccctaac tccgcccagt 3600
tccgcccatt ctccgccccca tggctgacta atttttttta tttatgcaga ggccgaggcc 3660
gcctcggcct ctgagctatt ccagaagtag tgaggaggct tttttggagg cctaggcttt 3720
tgcaaaaagc attgcagctt attgcagctt caaataaagc aatagcatca 3780
caaatttcac aaataaagca tttttttcac tgcattctag ttgtggtttg tccaaactca 3840
tcaatgtatc ttatcatgtc tggatccgct gcattaatga atcggccaac gcgcggggag 3900
aggcggtttg cgtattgggc gctcttccgc ttctctcgct actgactcgc tgcgctcggg 3960
cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggg tatccacaga 4020
atcaggggat aacgcaggaa agaacatgtg agcaaaaagg ccaggaaccg 4080
taaaaaggcc gcgttgctgg cgtttttcca taggctcgc cccctgacg agcatcacia 4140
aaatcgacgc tcaagtcaga ggtggcgaaa ccgcagagga ctataaagat accaggcgtt 4200
tccccctgga agctccctcg tgcgctctcc tgttccgacc ctgcccgtta ccggatacct 4260
gtccgccttg ctcccttcgg gaagcgtggc gctttctcaa tgctcacgct gtaggtatct 4320
cagttcggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc ccgttcagcc 4380
cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aacccggtta gacacgactt 4440
atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcggtgc 4500
tacagagttc ttgaagtggg ggcctaacta cggctacact agaaggacag tatttgggat 4560
ctgcgctctg ctgaagccag ttaccttcgg ggtagctctt gatccggcaa 4620
acaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa 4680
aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaacga 4740
aaactcacgt taagggattt tggatcatgag attatcaaaa aggatcttca cctagatcct 4800
tttaaaattaa aaatgaagtt ttaaatcaat ctaaaagtata tatgagtaaa cttggtctga 4860
cagttaccaaa tgcttaatac gtgaggcaac tatctcagcg atctgtctat ttcgttcac 4920
catagttgcc tgactccccg tcgtgtagat aactacgata cgggaggggt taccatctgg 4980
cccagtgct gcaatgatac cgcgagaccc acgctcaccc gctccagatt tatcagcaat 5040
aaaccagcca gccggaagg ggcagcgag aagtggctct gcaactttat ccgcctccat 5100
ccagctctatt aattgttgcc ggggaagctag agtaagtagt tcgccagtta atagtttgcg 5160
caacgtttgtt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc 5220
attcagctcc ggttcccaac gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa 5280
agcggttagc tccttcgggc ctccgatcgt tgtcagaagt aagttggccg cagtgttatc 5340
actcatgggt atggcagcac tgcataattc tcttactgtc atgccatccg taagatgctt 5400
ttctgtgact ggtgagtact caaccaagtc tagtgtatgc ggccgacgag 5460
ttgctcttgc ccggcgtaaa tacgggataa taccgcgcca catagcagaa ctttaaaagt 5520
gctcatcatt ggaaaacgtt ctccggggcg aaaactctca aggatcttac cgctgttgag 5580
atccagttcg atgtaaccca ctcgcgacc caactgatct tcagcatctt ttactttcac 5640
cagcgtttct gggtagcaaa aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc 5700
gacacggaaa tgttgaatac tcatactctt cctttttcaa tattattgaa gcatttatca 5760
gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaaatagg 5820
ggttccgcgc acatttcccc gaaaagtggc acctg 5855

```

<210> 128

<211> 303

<212> DNA

<213> Artificial Sequence

<220>

<223> Human FER-1 Promoter

<400> 128

```

tccatgacaa agcacttttt gagcccaagc ccagcctagc tcgagctaaa cgggacacaga 60
gacgccaccg ctgtcccaga ggcagtcggc taccgggtccc cgctcccag ctcccgcaga 120
gcgcgcgagg gcctccagcg gccgcccctc cccacagca ggggcggggg cccgcgccca 180
ccggaaggag cgggctcggg gcgggcggcg ctgattggcc ggggcggggc tgacgccgac 240
gcggctataa gagaccacaa gcgaccgcga gggccagacg ttcttcgccc agagtcgggt 300
acc 303

```

-103-

<210> 129
 <211> 6521
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pIRES-BSR Plasmid

<400> 129
 tcaatattgg ccattagcca tattattcat tgggtatata gcataaatca atattggcta 60
 ttggccattg catacgttgc atctatatca taatatgtac atttataattg gctcatgtcc 120
 aatatgaccg ccatgttggc attgattatt gactagttat taatagtaat caattacggg 180
 gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc 240
 gcctggctga ccgcccacg acccccggcc attgacgtca ataatgacgt atgttcccat 300
 agtaacgcca atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaacctgc 360
 ccacttggca gtacatcaag tgtatcatat gccaaagtcg cccctattg acgtcaatga 420
 cggtaaatgg ccgcctggc attatgccc gtacatgacc ttacgggact ttcctacttg 480
 gcagtacatc tacgtattag tcatcgctat taccatgggt atgcggtttt ggcagtacac 540
 caatgggcgt ggatagcggg ttgactcacg gggatttcca agtctccacc ccattgacgt 600
 caatggggag ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaacaactg 660
 cgatcgcccg ccccgttgac gcaaattggc ggtaggcgtg tacgggtggga ggtctatata 720
 agcagagctc gtttagtgaa ccgtcagatc actagaagct ttattgcggt agtttatcac 780
 agttaaatgg ctaacgcagt cagtgcctct gacacaacag tctcgaactt aagctgcagt 840
 gactctctta aggtagcctt gcagaagttg gtcgtgaggc actgggcagg taagtatcaa 900
 ggttacaaga caggtttaag gagaccaata gaaactgggc ttgtcgagac agagaagact 960
 cttgcgtttc tgataggcac ctattggctc tactgacatc cactttgcct ttctctccac 1020
 aggtgtccac tcccagttca attacagctc ttaaggctag agtacttaac acgactcact 1080
 ataggctagc ctcgagaatt cacgcgtcga gcatgcatct agggcggcca attccgcccc 1140
 tctccctccc cccccctaa cgttactggc cgaagccgct tgggaataagg ccggtgtgcg 1200
 tttgtctata tgtgattttc caccatattg ccgtcttttg gcaatgtgag ggcccggaaa 1260
 cctggccctg tcttcttgac gagcattcct aggggtcttt cccctctcgc caaaggaatg 1320
 caaggtctgt tgaatgtcgt gaaggaagca gttcctctgg aagcttcttg aagacaaaca 1380
 acgtctgtag cgaccctttg caggcagcgg aacccccac ctggcgacag gtgcctctgc 1440
 ggccaaaagc cagctgtata agatacacct gcaaaggcgg cacaacccca gtgccacgtt 1500
 gtgagttgga tagttgtgga aagagtcaaa tggctctcct caagcgtatt caacaagggg 1560
 ctgaaggatg ccagagggt accccattgt atgggatctg atctggggcc tcggtgcaca 1620
 tgctttacat gtgttttagtc gaggttaaaa aaacgtctag gcccccgaa ccacggggac 1680
 gtggttttcc tttgaaaaac acgatgataa gcttgccaca acccaccatg aaaacattta 1740
 acatttctca acaagatcta gaattagtag aagtagcgac agagaagatt acaatgcttt 1800
 atgaggataa taacatcat gtgggagcgg caattcgtag gaaaacagga gaaatcattt 1860
 cggcagtaca tattgaagcg tatataggac gagtaactgt ttgtgcagaa gccattgcga 1920
 ttggtagtgc agtttcgaat ggacaaaagg attttgacac gattgtagct gttagacacc 1980
 cttattctga cgaagtagat agaagtattc gagtggtaag tccttgtggt atgtgtaggg 2040
 agttgatttc agactatgca ccagattgtt ttgtgttaat agaaatgaat ggcaaggtag 2100
 tcaaaactac gattgaagaa ctcatccac tcaaataata ccgaaattaa aagttttacc 2160
 ataccaagct tggcggggcg ccgcttcctt ttagtgaggg ttaatgcttc gagcagacat 2220
 gataagatac attgatgagt ttggacaaac cacaactaga atgcagtga aaaaatgctt 2280
 tatttgtgaa atttgtgatg ctattgcttt atttghtaacc attataagct gcaataaaca 2340
 agttaacaac aacaattgca ttcattttat gtttcagggt cagggggaga tgtgggaggt 2400
 tttttaagc aagtaaaacc tctacaaatg tggtaaaatc cgataaggat cgatccgggc 2460
 tggcgtaata gcgaagagcg ccgcaccgat cgcccttccc aacagttgcg cagcctgaat 2520
 ggcgaaatga cgcgccctgt agcggcgcat taagcgcggc ggggtgtggtg gttacgcgca 2580
 gcgtgaccgc tacacttgcc agcgccttag cgcccgctcc tttcgctttc ttcccttccc 2640
 ttctcgccac gttcgccggc tttcccgtc ttcggggctc cctttagggt gatggttcac 2700
 tccgatttag agctttacgg cacctcgacc gcaaaaaact tgatttgggt gatggttcac 2760
 gtagtgggac atcgccctga tagacggttt ttcgcccctt gacgttggag tccacgttct 2820
 ttaatagtgg actctgttcc caaactggaa caaactcaa ccttatctcg gtctattctt 2880
 ttgatttata agggattttg ccgatttcgg cctattggtt aaaaaatgag ctgatttaac 2940
 aaatatttaa cgcgaatttt taacgtttac aatttcgcct gatgcggtat 3000
 tttctcctta cgcattctgt cggattttca caccgcatac gcggatctgc gcagcaccat 3060
 ggccgtgaaat aacctctgaa agaggaaact ggtaggtac cttctgaggc ggaaagaacc 3120
 agctgtggaa tgtgtgtcag ttagggtgtg gaaagtcccc aggtcctcca gcaggcagaa 3180
 gtagcgaag catgcatctc aattagttag caaccaggtg tggaaagtcc ccaggtccc 3240
 cagcaggcag aagtatgcaa agcatgcac tcaattagtc agcaaccata gtcccgcccc 3300
 taactccgcc catcccgccc ctaactccgc ccagttccgc cctattctcg ccccatggct 3360
 gactaatttt ttttatttat gcagaggccg aggcgcgctc ggccctctgag ctattccaga 3420
 agtagtgagg aggccttttt ggaggcctag gcttttgcaa aaagcttgat tcttctgaca 3480

-104-

caacagtctc	gaacttaagg	ctagagccac	catgattgaa	caagatggat	tgcacgcagg	3540
ttctccggcc	gcttgggtgg	agaggctatt	cggctatgac	tgggcacaa	agacaatcgg	3600
ctgctctgat	gccgccgtgt	tccggctgtc	agcgcagggg	cgcccgggtc	tttttgtcaa	3660
gaccgacctg	tccgggtgcc	tgaatgaact	gcaggacgag	gcagcgcggc	tatcgtggct	3720
ggccacgacg	ggcgttcctt	gcgcagctgt	gctcgacgtt	gtcaactgaag	cggggaaggga	3780
ctggctgcta	ttggggcgaag	tgcgggggca	ggatctcctg	tcatctcacc	ttgctcctgc	3840
cgagaaagta	tccatcatgg	ctgatgcaat	gcggcggctg	catacgcttg	atccggctac	3900
ctgcccattc	gaccaccaag	cgaaacatcg	catcgagcga	gcacgtactc	ggatggaagc	3960
cggctcttgc	gatcaggatg	atctggacga	agagcatcag	gggctcgcgc	cagccgaact	4020
gttcgccagg	ctcaaggcgc	gcatgcccga	cggcgaggat	ctcgtcgtga	cccatggcga	4080
tgcctgcttg	ccgaatatca	tgggtgaaaa	tggccgcctt	tctggattca	tgcactgtgg	4140
ccggctgggt	gtggcggacc	gctatcagga	catagcgttg	gctaccctgt	atattgctga	4200
agagcttggc	ggcgaatggg	ctgaccgctt	cctcgtgctt	tacggatatcg	ccgctcccga	4260
ttcgcagcgc	atcgcttctt	atcgcttctt	tgacgagttc	ttctgagcgg	gactctgggg	4320
ttcgaaatga	ccgaccaagc	gacgccccac	ctgccatcac	gatggccgca	ataaaatatc	4380
tttattttca	ttacatctgt	gtgttggttt	tttgtgtgaa	tcgatagcga	taaggatccg	4440
cgtatgggtc	actctcagta	caatctgctc	tgatgccgca	tagttaagcc	agccccgaca	4500
cccgccaaca	cccgtcgacg	cgccctgacg	ggcttgtctg	ctcccggcat	ccgcttacag	4560
acaagctgtg	accgtctccg	ggagctgcat	gtgtcagagg	ttttcaccgt	catcacccga	4620
acgcgcgaga	cgaaagggcc	tcgtgatacg	cctattttta	taggttaatg	tcatgataat	4680
aatggttttc	tagacgtcag	gtggcacttt	tcggggaaat	gtgcgcggaa	cccctatttg	4740
tttatttttc	taaatacatt	caaatatgta	tccgctcatg	agacaataac	cctgataaat	4800
gcttcaataa	tattgaaaaa	ggaagagtat	gagtattcaa	catttccgtg	tgcgcccttat	4860
tccctttttt	gcggtcattt	gccttccctgt	ttttgctcac	ccagaaacgc	tgggtgaaagt	4920
aaaagatgct	gaagatcagt	tgggtgcacg	agtgggttac	atcgaactgg	atctcaacag	4980
cggtaagatc	cttgagagtt	ttcgccccga	agaacgtttt	ccaatgatga	gcacttttaa	5040
agttctgcta	tgtggcgcg	tattatcccg	tattgacgcc	gggcaagagc	aactcggtcg	5100
ccgcatacac	tattctcaga	atgacttgg	tgagtactca	ccagtacacg	aaaagcatct	5160
tacggatggc	atgacagtaa	gagaattatg	cagtgtcgcc	ataaccatga	gtgataaac	5220
tgcggccaac	ttacttctga	caacgatcgg	aggaccgaag	gagctaaccg	ctttttttgca	5280
caacatggg	gatcatgtaa	ctcgccctga	tcgttgggaa	ccggagctga	atgaagccat	5340
accaaaccgac	gagcgtgaca	ccacgatgcc	tgtagcaatg	gcaacaacgt	tgcgcaaact	5400
attaactggc	gaactactta	ctctagcttc	ccggcaacaa	ttaatagact	ggatggaggc	5460
ggataaaagt	gcaggaccac	ttctgcgctc	ggcccttccg	gctggctgg	ttattgctga	5520
taaactctgga	gccgggtgag	gtgggtctcg	cggatcatt	gcagcactgg	ggccagatgg	5580
taagccctcc	cgtatcgtag	ttatctacac	gacggggagt	caggcaacta	tggatgaacg	5640
aaatagacag	atcgctgaga	taggtgcctc	actgattaag	cattggtaac	tgtcagacca	5700
agtttactca	tatatacttt	agattgat	aaaacttcat	ttttaattta	aaaggatcta	5760
ggtgaagatc	cttttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcgttcca	5820
ctgagcgtca	gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	ttttctgcg	5880
cgtaatctgc	tgcttgcaaa	caaaaaaac	accgctacca	gcggtgggtt	gtttgcccga	5940
tcaagagcta	ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	6000
tactgtcctt	ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcaccgcc	6060
tacatacctc	gctctgctaa	tctgttacc	agtggctgct	gccagtggcg	ataagtctgt	6120
tcttaccggg	ttggactcaa	gacgatagtt	accggataag	gcgcagcgg	cgggtgaa	6180
ggggggttcg	tgcacacagc	ccagcttgg	gcgaacgacc	tacaccgaac	tgagatacct	6240
acagcgtgag	ctatgagaaa	gcgccacgct	tcccgaagg	agaaaggcgg	acaggatcc	6300
ggtaagcggc	aggggtcggaa	caggagagcg	cacgaggag	cttccagggg	gaaacgcctg	6360
gtatctttat	agtccgtgctg	ggtttcgcc	cctctgactt	gagcgtcgat	ttttgtgatg	6420
ctcgtcaggg	gggcgggagcc	tatggaaaaa	cgccagcaac	gcggcctttt	tacggttcct	6480
ggccttttgc	tggccttttg	ctcacatggc	tgcacagatc	t		6521